AD-A080 407 AIR FORCE INST OF TECH WRIGHT-PATTERSON AFB OH SCHOO--ETC F/G 12/1

CROSS VALIDATION OF SELECTION OF VARIABLES IN MULTIPLE REGRESSI--ETC(U)

DEC 79 J R CAFARELLA

UNCLASSIFIED AFIT/GOR/MA/79D-2 NL

1 of 2

AD
A080407

(1)
B.S.

LEVEL II

DDC
RECEIVED
FEB 7 1980
A

UNITED STATES AIR FORCE
AIR UNIVERSITY
AIR FORCE INSTITUTE OF TECHNOLOGY
Wright-Patterson Air Force Base, Ohio

S DOCUMENT IS BEST QUALITY PRACTICABLE.
COPY FURNISHED TO DDC CONTAINED A
NIFICANT NUMBER OF PAGES WHICH DO NOT
RODUCE LEGIBLY.

DISTRIBUTION STATEMENT A
Approved for public release;
Distribution Unlimited

80 2 5 144

DISCLAIMER NOTICE

THIS DOCUMENT IS BEST QUALITY PRACTICABLE. THE COPY FURNISHED TO DDC CONTAINED A SIGNIFICANT NUMBER OF PAGES WHICH DO NOT REPRODUCE LEGIBLY.

CROSS VALIDATION OF
SELECTION OF VARIABLES
IN MULTIPLE REGRESSION

THESIS

GOR

AFIT/[redacted]/MA/79D-2

Joseph R. Cafarella, Jr.
2 Lt USAF

Approved for public release; distribution unlimited.

AFIT/GOR/MA/79D-2

CROSS VALIDATION OF SELECTION OF VARIABLES IN MULTIPLE REGRESSION

THESIS

Presented to the Faculty of the School of Engineering

of the Air Force Institute of Technology

Air University

In Partial Fulfillment of the

Requirements for the Degree of

Master of Science

by

Joseph R. Cafarella, Jr

2 Lt. USAF

Graduate Operations Research

Dec[redacted]79

Approved for public release; distribution unlimited

Table of Contents

Page

Acknowledgements

In any research effort, success or failure depends on the efforts of many individuals. With this is mind, I would like to express my sincere gratitude to Dr. David R. Barr for his continued support, expertise, and guidance throughout this research.

I would like to thank Lt. Colonel Charles W. McNichols who assisted me in various computer problems I encountered, and Lt. Colonel Saul Young who acted as my reader on my thesis committee.

I would also like to thank Diane Summers and Dan Ferins of the Systems Evaluation Branch (AAA-3) of the AF Avionics Laboratory for supplying the necessary data and background information used in this research.

Abstract

Techniques and criterion for selection of the "best" subset of variables to be used in a regression model are reviewed.

A model was developed using the Automatic Interaction Detection (AID) algorithm as a pre-screening device for locating those variables most important to the regression including interaction terms.

Five previous models including the one developed by AID and one developed by Westinghouse on avionic characteristic data are used in cross validation experiments to determine the predictive power of these models on a new set of data points using the same set of variables. A cross validation R^2 value is discussed as a criterion for choosing between competing models.

List of Figures

List of Tables

CROSS VALIDATION OF SELECTION OF

VARIABLES IN MULTIPLE REGRESSION

I Introduction

Background

Long term DoD planning goals require than operational and support costs on all projects be reduced. Managers of these projects are challenged by the need for accurate evaluation of these projects in the early design stages. A question arises, however, concerning whether model development and enhancement should be contracted out-of-house or done using available efforts of Air Force personnel in-house. Performing a cost analysis in-house would surely reduce costs. Also, performing an in-house cost analysis would benefit the user of the model by providing first hand knowledge of the impacts of updates and changes in the data base on the final results and may discover intermediate results unknown to a contractor.

One prerequisite for the user to perform in-house analysis is the availability of the necessary computer packages. Another is the knowledge of the user in applying other effective methods of analyzing the goodness of fit of the models other than the R^2 value or F-statistic discussed in the next chapter. Once the user of the model attains these prerequisites, in-house analysis can be performed.

Since these prerequisites for an in-house capability of cost estimation were not available at the time, the Systems Evaluation Branch (AAA-3) of the Air Force Avionics Laboratory at Wright-Patterson Air Force Base requested that the Westinghouse Electric Corporation perform a regression analysis on certain characteristics of Line

Replaceable Avionic Units (LRUs).

The Westinghouse approach was to select "candidate" LRUs for inclusion in the data base, collect data on design and logistic characteristics on the LRUs, perform a regression analysis on the data, then use the resulting cost and parametric relationships to construct a model. The resulting model was named the Avionics Laboratory Predictive Operations and Support (ALPOS) model [36].

One of the problems Westinghouse encountered, which most analysts encounter also, involved the process used in the selection of the data. Probably the most important element in the research is the nature of the data which was used. Many different situations can arise from "bad" data and wrong assumptions about the data such as whether the data subset collected is statistically different from the underlying population or whether multicolinearity exists between variables.

In the initial phase, several LRUs were identified and considered for inclusion in the data base from a wide variety of avionic units placed on various types of aircraft. The LRU selection was naturally constrained by the availability of the data and on the number of aircraft on which the LRU was installed. This initial data base (Phase I) consisted of sixty-three LRUs from seven different aircraft.

For their regression analysis, Westinghouse used the Linear Least-Squares Curve Fitting Program (LLSCFP) developed by Daniel and Wood [8]. This computer program uses over thirty statistics and five types of plots in assisting the analyst develop meaningful variable relationships.

In his Masters thesis, Captain Larry Pulcher attempted to provide the means for the members of AAA-3 to conduct their own in-house cost-estimation analysis by developing and testing criterion for selection of variables in a regression analysis including iterative techniques using the Statistical Package for the Social Sciences (SPSS), all possible regressions using the International Mathematical Statistical Library (IMSL) routine RLEAP, and the Omnitab computer package used to compute prediction intervals.

Both Westinghouse and Pulcher had available a set of potential variables which could be considered for inclusion in the model, however, both sets of variables were too large (more variables than data points). Westinghouse used an approach in which "candidate" variables were screened and tested before admission to the model. Pulcher used a screening technique to eliminate certain candidate variables before hand.

Focus of this Research

Westinghouse has recently updated the data collected in the initial phase. This new Phase II data base includes sixty-five additional LRUs plus six previous ones placed on different aircraft for a total of seventy-one LRUs. Also, four additional aircraft have been included. See Table I for a summary of the LRUs investigated.

One objective of this research is to review past research in the area of selection of variables in a regression analysis in the hope of stimulating thoughts and ideas of those analysts interested in combining talents on this subject.

A second objective of this research is to examine the three previous models developed by Pulcher and the Phase I model developed by Westinghouse and determine which of the models predicts the Phase II data the best.

A third objective of this research is to use the Automatic Interaction Detection (AID) algorithm documented by Sonquist and Morgan [33, 34] to prescreen variables from the entire data set and create a model based on the Phase I data and perform the same predictive tests mentioned above using the Phase II data. A Leaps and Bounds algorithm was used to assess various AID models to determine which one should be represented in the subsequent analysis.

Finally, updated coefficients were calculated for the best predictive model determined in objectives two and three above.

TABLE I

Summary of LRUs Investigated

Aircraft	PHASE I	PHASE II	TOTAL
F4E	11	3	14
RF4C	8	-	8
F15A	10	20	30
B52G/H	18	1	19
KC135A	5	-	5
C130E	5	6	11
C5A	6	9	15
F106A	-	2	2
F111D	-	20	20
FB111A	-	10	10
TOTAL	63	71	134

II Concept Overview

Theory of Least Squares Regression

The fundamental premise of a regression analysis is to build a model useful in predicting a single dependent or criterion variable from a set of independent or predictor variables. There are many different types of models which can be created such as general linear discussed in the following section, non linear, logarithmic, polynomial, reciprocal and multiplicative. This research deals mainly with linear, polynomial and logarithmic models.

Assumptions

Before any statistical inferences can be made and tests performed on the significance of the coefficient estimates and the independent variable, certain assumptions must be made about the data and about the probability distribution of the random error.

The first assumption is that the data is a sample from the target population. The second assumption is that the random variable ε, the error term, is:

(1) statistically independent

(2) identically distributed

(3) from a population with zero mean

(4) normally distributed

In other words, $\varepsilon \sim N(0, \sigma^2)$ which means that ε is from a normal probability density function with a mean of zero and a variance of σ^2. Also, since nothing is known about the probability distributions describing these error terms, the Central Limit Theorem guarantees that

if we can assume independence, then the sum will tend to be normally distributed. Also, if we can assume that all the error terms have identical probability distributions, then we insure that each of them have the same variance.

Method of Least Squares

The general form of the linear least squares model is

$$Y = \beta_o + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_j X_j + \cdots + \beta_k X_k + \varepsilon \qquad (1)$$

where Y is the observed value of the dependent variable

X_j is the observed value of the jth independent variable

β_o is the constant term

β_j is the regression coefficient for the jth independent variable

ε is the random variable accounting for the error

k is the number of independent variables

Note that X_j can be the transformation of an original observation. For example, the Product of Powers model

$$Y = \beta_o \, X_1^{\beta_1} X_2^{\beta_2} \qquad (2)$$

can be transformed in a linear sense to

$$\ln(Y) = \beta_o + \beta_1 \ln(X_1) + \beta_2 \ln(X_2) \qquad (3)$$

or

$$Y^* = \beta_o + \beta_1 X_1^* + \beta_2 X_2^* \qquad (4)$$

where the "*" indicates the transformed variable in equation (4).

If there are n dependent variables, equation (1) can be written:

$$Y_i = \beta_o + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_j X_{ij} + \ldots + \beta_k X_{ik} + \varepsilon_i \qquad (5)$$

where $i = 1, 2, \ldots, n$

Since it is very difficult to discuss the multiple regression case in algebraic terms, matrix notation will be used. Equation (5) can be written as:

$$\underline{Y} = X\underline{\beta} + \underline{\varepsilon} \tag{6}$$

where $\underline{Y}$ represents an n-element column vector of observed values of the dependent variable:

$$\underline{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} \tag{7}$$

X represents an n x K + 1 matrix. The first column contains all ones representing the constant term. The other columns represent the X_{ij} values:

$$X = \begin{bmatrix} 1 & X_{11} & X_{12} & \cdots & X_{1K} \\ 1 & X_{21} & X_{22} & \cdots & X_{2K} \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ 1 & X_{n1} & X_{n2} & \cdots & X_{nK} \end{bmatrix} \tag{8}$$

$\underline{\beta}$ represents a K + 1 x 1 column vector of regression coefficients:

$$\underline{\beta} = \begin{bmatrix} \beta_o \\ \beta_1 \\ \beta_2 \\ \cdot \\ \cdot \\ \cdot \\ \beta_k \end{bmatrix} \tag{9}$$

$\underline{\varepsilon}$ represents the n-element column vector of error terms:

$$\underline{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \cdot \\ \cdot \\ \cdot \\ \varepsilon_n \end{bmatrix} \quad (10)$$

The objective of the least squares technique is to fit a line through a set of data points so that the sume of the squared differences between Y_i ($i = 1, 2, \ldots, n$), the actual values of the dependent variable, and $\hat{Y}_i$, the estimated value of the dependent variable, is minimized.

$\hat{Y}$ is defined algebraically as:

$$\hat{Y}_i = \beta_o + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_j X_{ij} + \ldots + \beta_k X_k \quad (11)$$

or in matrix notation as:

$$\hat{\underline{Y}} = X\underline{\beta} \quad (12)$$

The random error term ε *is the difference between* $\underline{Y}$ *and* $\hat{\underline{Y}}$ and can be written as follows:

$$\underline{\varepsilon} = \underline{Y} - \hat{\underline{Y}} \quad (13)$$

A two-dimensional graphical depiction of a regression line using three data points is shown in Figure 1.

The ideal situation is to have each of the error terms equal to zero. That way, the regression model would fit the data points exactly. In most cases, however, this is not possible so minimizing the sum of the error terms is the best solution. In order to keep the mathematics relatively easy, the error terms are made positive by squaring each term

before summation. This sum of squared errors (SSE) can be written as:

$$SSE = \sum_{i=1}^{n} (\varepsilon_i)^2 = \underline{\varepsilon}' \ \underline{\varepsilon} \tag{14}$$

where $\underline{\varepsilon}'$ is the transposed matrix $\underline{\varepsilon}$. The objective can now be stated as follows:

Find $\underline{\beta}$ to minimize:

$$SSE = \underline{\varepsilon}' \ \underline{\varepsilon} = (\underline{Y} - \hat{\underline{Y}})' \ (\underline{Y} - \hat{\underline{Y}}) = (\underline{Y} - X\underline{\beta})' \ (\underline{Y} - X\underline{\beta}) \tag{15}$$

Using a straightforward application of Lagrange's Multipliers on equation (15), one estimator of $\underline{\beta}$ which minimizes SSE is:

$$\hat{\underline{\beta}} = (X'X)^{-1} \ X'\underline{Y} \tag{16}$$

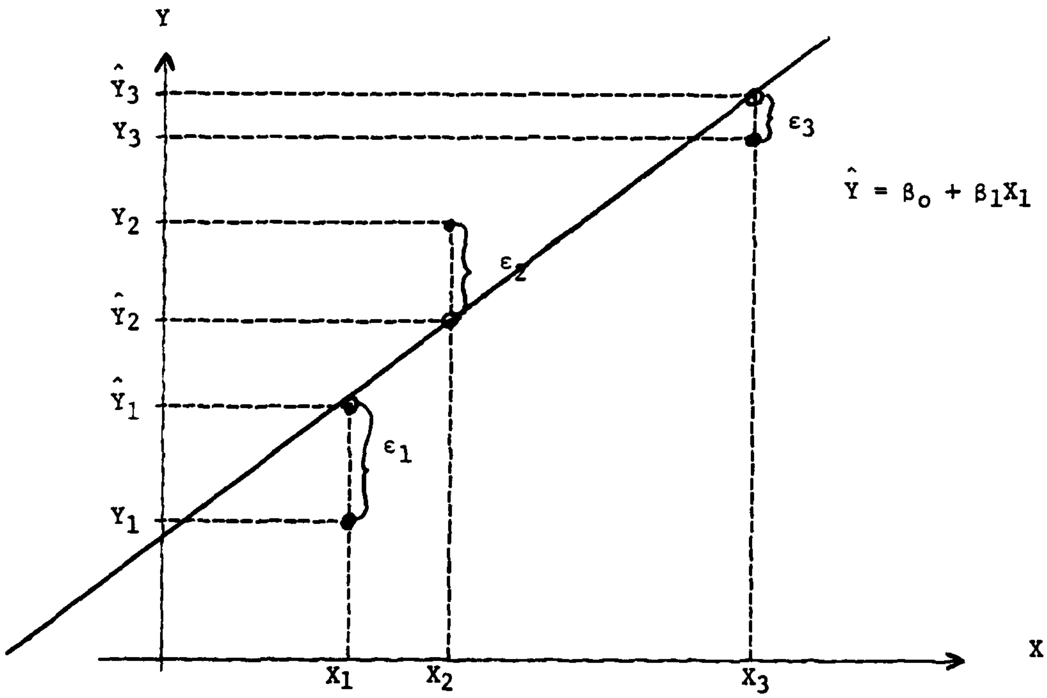


FIGURE 1 Regression Line

It is known, however, that a regression model containing these estimates of $\underline{\beta}$ will not explain all of the variability in the dependent variable $\underline{Y}$. Some of the variability in $\underline{Y}$ will be explained by the regression model and the remaining portion is left unexplained. This idea can be stated as follows:

$$SST = SSR + SSE \tag{17}$$

where SST is the total sum-of-squares or the total variability in the dependent variable and is defined as:

$$SST = \sum_{i=1}^{n} (Y_i - \bar{Y})^2 = \sum_{i=1}^{n} Y_i^2 - n\bar{Y} \tag{18}$$

or

$$SST = \underline{Y}' Y - n \bar{Y}^2 \tag{19}$$

SSR is the regression sum-of-squares or the variability in the dependent variable explained by the regression model and is defined as:

$$SSR = \sum_{i=1}^{n} (\hat{Y}_i - \bar{Y})^2 \tag{20}$$

or

$$SSR = \underline{\hat{\beta}}' X' \underline{Y} - n\bar{Y}^2 \tag{21}$$

SSE is the residual or error sum-of-squares or the remaining amount of variability which is left unexplained and is defined by equation (15).

<u>Measures of Merit:</u>

Since SST depends only on the values of the dependent variables, Y_i, it is constant for any given set of n observations. Also, since SSE is being minimized, this makes SSR as large as possible. It is then reasonable to assume that the ratio of SSR to SST would be an adequate indicator of the goodness of fit of the model to the data

and a good measure of merit of the regression. This ratio is denoted as R_y^2, or simple as R^2, and is called the coefficient of determination or the multiple R-squared value.

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} \qquad 0 \leq R^2 \leq 1 \tag{22}$$

According to Theil [35:178], the sample value of R^2 is somewhat biased due to the degrees of freedom used in its calculation. Theil suggests that a better measure of merit is $\bar{R}^2$, defined as the adjusted multiple correlation coefficient.

$$\bar{R}^2 = 1 - (1-R^2)\left(\frac{n-1}{n-k}\right) \tag{23}$$

or

$$\bar{R}^2 = 1 - (1-R^2)\left(\frac{n-1}{n-k-1}\right) \tag{24}$$

if a constant term is included in the model, or equivalently as

$$\bar{R}^2 = R^2 - (1-R^2)\left(\frac{k-1}{n-k-1}\right) \tag{25}$$

In either of the cases above, $\bar{R}^2$ is always less than or equal to R^2. It must be noted, however, that $\bar{R}^2$ is not an unbiased estimator, though it still has some merit because when the number of variables being estimated, k, becomes large compared to the number of observations or data points, n, it still gives an optimistic picture of the amount of variability in the dependent variable explained by the regression model.

$\bar{R}^2$ can also be defined as:

$$\bar{R}^2 = 1 - \frac{MSE}{MST} \tag{26}$$

where MSE, mean square error $= \frac{SSE}{n-k-1}$

and MST, mean square total $= \frac{SST}{n-1}$

Thus, MSE = MST* $(1-\bar{R}^2)$, and minimizing MSE maximizes $\bar{R}^2$.

Mosier [26] has suggested a measure of merit similar to R^2 which measures the predicting power of a model. Based on a model using the original set of old data (Phase I), the estimated value for each data point, $\hat{Y}_i$, was calculated. The cross validation SSE (c.v. SSE) was then calculated by the following equation: c.v. SSE $= \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2$, where the Y_is are the actual (observed) values from the new set of data (Phase II). Notice that the c.v. SSE is not the same as SSE because both Y_i and $\hat{Y}_i$ did not come from the same sample.

The c.v. SSE is then used to calculate the cross validation R^2 by c.v. $R^2 = 1 - \frac{\text{c.v. SSE}}{\text{SST}}$. Here, c.v. R^2 indicates the predictive power of the old models on the new data.

III Review of Past Research

There is a considerable amount of literature examining the many efforts that have been made to determine the "best" subset of independent variables that should be included in a regression model so that the amount of unexplained variance in the dependent variable is reduced. Many criteria for selection of these variable subsets have been examined, yet no one best criterion has been found.

Draper and Smith [10:163] point out two conflicting viewpoints on this subject. At one extreme, all variables could be included in the model for predictive purposes, however, though the values predicted may be reliable, as the number of variables in the model approaches the number of data points or observations, R^2 will naturally become close to one, thus implying a false sense of importance of the model to the unexperienced analyst.

At the other extreme, the model could include as few variables as possible so that the predictions are still reliable and the costs of maintaining and updating the data base is kept at a minimum. A compromise between these two viewpoints is suggested and is considered to be the "best" approach.

One would like to examine all of the 2^k possible regressions of the dependent variable in the search for the best equation, however, not only would there be computational and time limitations on the computer which make this approach impractical, but there is the remaining problem of specifically defining what is meant by the "best" regression model and when it has been found. This chapter reviews some of the research that has been done in this subject area.

Probably the most well known research on the subject of variable selection and regression analysis is that of Draper and Smith. Four different regression approaches have been devised including all possible regressions, backward elimination, forward selection, and stepwise regression.

In the All Possible Regressions technique, all 2^k possible regressions are considered. Thus a ten variable model would require the examination of 2^{10} or 1024 possible regressions. Each model is ordered by some criterion such as R^2 or $\bar{R}^2$ and compared. Often for large data bases, it becomes necessary to compute the residual mean square error and assess its magnitude to determine the best cut-off point for the total number of variables in the regression.

Recent research by analysts such as Schatzoff, Tsao, and Fienbert [31] have been able to reduce the number of calculations required from an order of k^3 to k^2, thus making this technique more practical, yet still relatively expensive to use. However, if the number of variables was reduced by methods such as the Chow test developed by Gregory Chow [7], this method becomes even more practical.

In the backward elimination method, a regression equation containing all possible variables is used as a starting point. A partial-F value is calculated for each variable and if a value is less than some specified tabular value, then that variable is removed from the model. Once a variable is removed from the model it is not susceptible to further consideration. A new regression is then computed and the process continues until no more variables can be eliminated from the model. Although this method is not thought of as the most powerful

methods to use in determining the best regression equation, Mantel [23] supports the method and points out its many advantages.

The forward selection process operates in a reverse manner from the backward elimination procedure. Variables enter the model one at a time until a model has been satisfied. Initially, partial-F statistics and partial correlation coefficients are calculated between each independent variable and the dependent variable. The variable most highly correlated will enter the regression equation. A new regression equation is then calculated and the process continues. Once a variable has entered the regression equation, there is no chance that it will be removed. This, however, is one of its faults. There is no attempt to determine the effect an entering variable has on the existing variables in the model.

In the stepwise regression procedure, however, an examination is made at each stage of inclusion of variables in the model to determine whether any variable or set of variables introduced previously lose their significance due to the introduction of a new variable. Thus, a variable which entered at an earlier stage,yet has been found unimportant due to the inclusion of a new variable,will be detected and removed from the model. For this reason, the stepwise procedure has been determined to be the most powerful regression technique.

In discussing various regression procedures, there are three important points that need mentioning. The first point is that the order of inclusion of the variables in the model is irrelevent. Thus, a variable which entered early in the model does not mean that it is more important than a variable which entered later. The second point

is that there is no guarantee that any of the previous methods will arrive at the best regression model. The third point is that there is also no guarantee that each of the previous methods will arrive at the same model or subset of variables. This is true between any set of regression procedures.

There are many more criterion for selecting variable subsets other than R^2 or the partial-F statistic. The remaining portion of this chapter is dedicated to mentioning those various research efforts.

Aitken [1] discusses the use of the Mean Square Prediction Error (MSPE) as a criterion for selecting variable subsets if the regression equation is used for prediction purposes rather than description purposes. In the later case, he prefers the use of the conventional R^2 value as a criterion. Allen [2] also discusses the use of the MSPE for selecting variable subsets.

The MSPE is defined as the expected value of the squared difference between the actual value of the independent variable, Y, and the estimated value, $\hat{Y}$. If all dependent variables are used in the regression equation, Aitkin defines the MSPE as follows:

$$\text{MSPE} = E[Y - \hat{Y}]^2 = \sigma^2 [\frac{n+1}{n} + (\underline{x} - \bar{\underline{x}})' S_{xx} (\underline{x} - \bar{\underline{x}})] \qquad (27)$$

where $\underline{x}$ is a row rector of X, $\bar{\underline{x}}$ is the vector of means, and S_{xx} is the matrix of cross products of the k independent variables: $S_{xx} = X'X$.

Allen defines the MSPE as follows:

$$\text{MSPE} = E[Y - \hat{Y}] = \sigma^2 + \text{Var}(\hat{Y}) + [E(\hat{Y}) + \underline{X}\ \underline{\beta}]^2 \qquad (28)$$

where the last term is the squared bias of prediction and the last two terms together are the Mean Square Error (MSE) of $\hat{Y}$.

Since the least squares predictor $\hat{Y}$ is unbiased, its variance is $\underline{x}(X'X)^{-1}\underline{x}\,\sigma^2$. If the last term is dropped, one gets:

$$MSPE_r = \sigma^2 + \underline{x}\,(X'X)^{-1}\underline{x}\,\sigma^2 \tag{29}$$

which Allen uses for the comparison of other predictors.

Kennedy and Bancroft [22] discuss using the average value of the MSPE over their sample as a criterion:

$$MSPE^* = \frac{1}{n}\sum_{i=1}^{n} \sigma^2 \left[\frac{n+1}{n} + (X_i - \bar{x})' S_{xx}^{-} \; (X_i - \bar{x}) \right] \tag{30}$$

$$= \frac{\sigma^2}{n}\,(n + k - 1)$$

where X has been assumed to follow a uniform distribution. Aitken, however, believed it more realistic to assume that all X values were independently and identically distributed. In either case, the objective is to chose the variable subset which minimizes the MSPE. If the subset of variables to be tested is specified in advance or simply fixed, the testing hypothesis becomes:

$$\begin{aligned} H_o &: MSPE - MSPE^1 \geq 0 \\ H_a &: MSPE - MSPE < 0 \end{aligned} \tag{31}$$

where $MSPE^1$ is the MSPE of the variable subset. If the null hypothesis, H_o, is not rejected, this means that the subset of variables is not statistically different from that of the total set of data and the subset may be considered for use in a prediction equation. A non-central F-statistic and test have also been developed by Aitken to estimate (31) depending on the assumed distribution and selection process of the independent variables. In the cases where the variable subsets are unknown, a simultaneous procedure, similar to the forward selection process developed by Draper and Smith, was developed by Garland [15]. In this procedure, variable subsets are chosen based on a central-F approximation to the multiple correlation coefficient.

Helms [16] discusses the use of the Average Estimated Variance (AEV) as a criterion for comparing competing linear models and explains why the Integrated Mean Square Error (IMSE) used as a criterion is not very useful in practice. The technique includes the computation of the AEV for each possible regression and the implementation of a stepwise procedure using the AEV as a criterion rather than R^2 or Mallows' C_p statistic. One advantage of the AEV has over R^2 and C_p is that it automatically incorporates information about the tradeoff between bias and variance when one enters or deletes variables in the model.

Furnival and Wilson [13] discuss a technique for computing the error sum of squares (SSE) for all possible regressions with minimal amount of calculations, and show how it is implemented in a branch and bound technique which they refer to as the Leaps and Bounds technique. This technique is useful in determining the best subset, and without examining all the possible subsets of variables.

The fundamental principal upon which their research is based is that $SSE(A) \leq SSE(B)$ where A is any set of independent variables and B is a subset of A. In other words, it is impossible for any subset of A to have a lower error sum of squares than A. Because of this, SSE(A) can be used as a lower bound in the analysis which means that subsets of A can be ignored in the search for the best given numbered variable subset.

In their technique, two search variations are described: horizontal and vertical. The horizontal variation explains regressions in a probability tree form and in a conventional or natural order so

that all one variable regressions, two variable regressions, etc. are easily observable. These regression trees are formed by beginning with all k variables in a regression and branching out on all possible k-1 variable subsets. The value of SSE is computed for each of the subsets and the subset with the smallest value will be the "best" k-1 variable subset. That subset will not be divided further as it provides a minimum value for that branch. Branching occurs elsewhere in the same manner as above until the best possible k-2, k-3, ..., 1 variable subsets are chosen.

Criterion for selecting these variable subsets is based on either R^2, $\bar{R}^2$, or Mallows C_p statistic. In a similar fashion, Narula and Wellington [25] introduce a branch and bound algorithm using the Minimum Sum of Weighted Absolute Errors (MSWAE) as a criterion for selecting variable subsets and involves the use of linear programming to minimize the sum of the absolute values of the residuals subjected to several constraints.

Andrews [4] discusses the use of regression and model building by medians and also introduces a robust method of analyzing data assumed not to have a Gaussian distribution with errors of equal variances.

Webster, Gunst, and Mason [37] discuss a modified least squares estimation procedure using latent roots and latent vectors of the correlation matrix of the dependent and independent variables. This has been found to be very useful when the matrix of independent variables is nearly singular.

In a more recent article, Park [29] discusses a strategy for selecting subsets of variables from a given linear mixture model developed by Scheffe [32], and applies the MSE as a criteria for screening the variables for model reduction.

In another recent article, Ellerton [11] investigates a method of applying linear programming to determine whether a given subset of variables is adequate in a regression model.

Surprisingly enough, very little cross-communication has been done concerning this very important subject, and I believe a joint analytical effort should be made testing these various criteria against various data bases in order to determine if there is one best method or criterion useful in predicting variable subsets to be used in a regression model.

IV Model Development and Selection of Variables

The Westinghouse Data Base

Senior engineers from Westinghouse collected most of the data in both Phase I and Phase II from on site visits to the Pentagon, AFLC Headquarters, ATC Headquarters, four Air Logistic Centers (ALCs), and several Air Force bases. While on site, interviews were conducted with technicians to verify the appropriateness of the LRUs originally selected and to identify possible alternatives.

At the completion of the Phase II data collection, the resulting data base contained 134 LRUs (See Appendix A), and thirty-three elements (variables plus indicators per LRU) (see Table II). After various variable transformations and modifications, twenty variables remained.

The first set of variables describe the aircraft type and avionics area and are indicators (zero or one). Three aircraft types including fighter, bomber and cargo and three avionic areas including sensory, communication and navigation were initially coded as follows:

Bomber	1	0
Cargo	0	1
Fighter	0	0

Sensory	1	0
Communication	0	1
Navigation	0	0

After additional investigation, the following set of indicator

TABLE II

Westinghouse Data Base Elements

1.	Bomber indicator variable (1 indicates Bomber aircraft)
2.	Cargo indicator variable (1 indicates Cargo aircraft)
3.	Sensory indicator variable (1 indicates sensory avaionics)
4.	Communications indicator variable (1 indicates comm avionics)
5.	Unit Price
6.	Volume (in^3)
7.	Weight (lbs)
8.	Component Count
9.	Percentage Digital Components
10.	Percentage Analog Components
11.	Percentage Electro-Mechanical Components
12.	Percentage Power Supply Components
13.	Percentage Transmitter Components
14.	Percentage Solid State Components
15.	Power Dissipation (watts)
16.	Utilization Factor (Operating hours/flying hour)
17.	Percentage Failures Detected by Automatic Test (BIT/FIT FACTOR)
18.	Number of Integrated Circuits
19.	Number of SRUs in the LRU
20.	Mean Time (flight hours) Between Failures
21.	Mean Time (flight hours) Between Maintenance Actions
22.	Maintenance Manhours - Scheduled (Organizational)
23.	Maintenance Manhours - Unscheduled (Organizational)
24.	Maintenance Manhours - Shop (Intermediate)
25.	Logistic Support Cost - Field
26.	Logistic Support Cost - Special Repair Center (Depot)
27.	Logistic Support Cost - Packaging and Transportatiou
28.	Logistic Support Cost - Condemnation Replenishments
29.	Training Costs
30.	Percentage LRUs Not Repairable This Station (%NRTS)
31.	Flying Hours (FH) (to normalize MMH and LSC)
32.	Percentage Condemned LRUs
33.	Specialized Repair Activity (Depot) Costs
34.	Quantity per Assembly
35.	Flying hours (to normalize Training costs)

variables was used in the regression analysis to denote interactions between aircraft type and avionics area:

LRUs in fighter aircraft navigation systems

LRUs in fighter aircraft sensory systems

LRUs in fighter aircraft communication systems

LRUs in bomber aircraft navigation systems

LRUs in bomber aircraft sensory systems

LRUs in bomber aircraft communication systems

LRUs in cargo aircraft navigation systems

LRUs in cargo aircraft communication systems

LRUs in cargo aircraft sensory systems were not included. The above set of indicators is coded as follows:

Fighter-Navigation	1	0	0	0	0	0	0
Bomber-Navigation	0	1	0	0	0	0	0
Cargo-Navigation	0	0	1	0	0	0	0
Fighter-Sensory	0	0	0	1	0	0	0
Bomber-Sensory	0	0	0	0	1	0	0
Fighter-Communication	0	0	0	0	0	1	0
Cargo-Communications	0	0	0	0	0	0	0

The next four independent variables are measures of physical characteristics. The Unit Price is measured in 1976 dollars per LRU and ranges in value from $153 to $220,943. The Volume is measured in cubic inches and ranges in value from 30 to 8200. The Weight is measured in pounds and ranges in value from one pound to 8200 pounds. Component Count is the number of electronic components and ranges in value from none to 7638.

The next five independent variables are categories of the different component types including Digital, Analog, Electromechanical, Power Supplies, and Transmitter, and are measured as a percentage of the total number of components having that characteristic. All values range from zero to 100 percent.

The next independent variables, Fraction Solid State, and the number of Integrated Circuits in each LRU are measures of LRU technology, the later ranging in value from zero to 4625.

The sixteenth independent variable is a measurement the Power Dissipation and is defined as the input power minus the transmit power, and ranges in value from six to 1640 watts.

The next independent variable represents a percentage of failures in LRUs detected by the Built-In-Test/Fault-Isolation-Test (BIT/FIT).

The last two independent variables are the Specialized Activity (Depot) Costs and the Quantity Per Assembly.

Westinghouse also identified several dependent variables. These include the Mean Time Between Failures (MTBF), the Mean Time Between Maintenance Actions (MTBMA), the Total Maintenance Man Hours per Operating Hour (MMH-UNS/OH), the Maintenance Man Hours in the Shop per Operating Hour (MMH-SHOP/OH), the Total Logistic Support Costs per Operating Hour (LSC-TOT/OH), the Field Logistic Support Cost per Operating Hour (LSC-FLD/OH), the Training Costs per Operating Hour (TRAIN/OH), and the percentage of LRUs not repairable this station (NRTS).

Only one of the dependent variables mentioned above will be used

in the analysis; LSC-TOT/OH. A list of all the variables used in this report and previous reports is contained in Table III.

Previous Models

In this section, five previous models (two developed by Westinghouse and three developed by Pulcher) are discussed.

The first Westinghouse model (Table IV) was based on the Phase I data and second (Table V) was based on the Phase II data. All variables in the first model are in linear form, quadratic form or logarithmic form.

The three models developed by Pulcher are described in Table VI and Table VII. Initially, Pulcher was able to create ninety-seven variables from the Product of Powers model of the form:

$$\ln Y = \alpha_o + \sum_{i=1}^{13} \alpha_i D_i + \sum_{j=1}^{6} \beta_{jo} \ln x_j + \sum_{j=1}^{6} \sum_{i=1}^{13} \beta_{ji} D_i \ln X_j \quad (31)$$

The D_i are indicator variables, and their function is to allow for coefficients to be different for subpopulations. For a simplified example, suppose we had:

$$\ln Y = \alpha_o + \alpha_1 D_1 + \beta_1 \ln X_1 + \beta_{11} D_1 \ln X_1 \quad (32)$$

For the subpopulation for which $D_1 = 0$, the model is:

$$\ln Y = \alpha_o + \beta_1 \ln X_1 \quad (33A)$$

while for the subpopulation for which $D_1 = 1$, the model is:

$$\ln Y = (\alpha_o + \alpha_1) + (\beta_1 + \beta_{11}) \ln X_1 \quad (33B)$$

Since there were only 63 data points, a method was needed to reduce the number of variables. Pulcher chose the Chow Test (also called the Test of Equality Between Subsets of Coefficients in Two Regressions),

TABLE III

List of Variables - Abbreviations

Name	Westinghouse	Pulcher	This Report
Bomber	IBOM	*	BOMBER
Cargo	ICAR	*	CARGO
Sensory	ISEN	*	SENSORY
Communication	ICOM	*	COMM
Navigation-Fighter	*	*	FGTNAV
Navigation-Bomber	IBMNAV	*	BOMNAV
Navigation-Cargo	*	*	CARNAV
Sensory-Fighter	*	SF	FGTSEN
Sensory - Bomber	*	SB	BOMSEN
Communication - Fighter	IFGCOM	CF	FGTCOM
Communication - Bomber	IBMCOM	CB	BOMCOM
Communication - Cargo	*	COMMC	CARCOM
Unit Price	UP	UP	UP
Volume	V	V	V
Weight	W	W	W
Component Count	CC	CC	CC
Component Density	CD	*	*
Power Dissipation	PD	PD	PD
Fraction Solid State	FSS	% SS	SS
Fraction Digital	FDI	% DIG	DIG
Fraction Analog	FAN	% AN	AN
Fraction Electromechanical	FEM	% EM	EM
Fraction Power Supply	FPS	% PS	PS
Fraction Transmitter	RXR	% XMTR	XMTR
Fraction BIT/FIT	BIT/FIT	BF	BITFIT
Number of Integrated Circuits	IC	*	IC
Specialized Repair Activity Costs	SRA	*	SRU
Quantity Per Assembly	QPA	*	QPA
Logistic Support Cost/ Operating Hour	LSC/OH	LSC/OH	LSC/OH
Maintenance Manhours/ Operating Hour	MMH/OH	*	*
Mean Time Between Failures	MTBF	*	*
Mean Time Between Maintenance Actions	MTBMA	*	*
Training Cost/Operating Hour	TRAIN/OH	*	*
Not Repairable This Station	NRTS	*	*

* Not used in the analysis

TABLE IV

Westinghouse Model - Phase I Data

$$\ln\ (LSC/OH) = b_o + \sum_{i=1}^{21} b_i X_i$$

$\bar{R}^2$ = .8916 R^2 = .9283 F-value - 25.3

i	b_i	X_i	Partial-F
0	-8.15108		
1	3.86111	(IBOM-.2857142857)	36.0
2	3,66533	(ICAR-.2698412698)	31.4
3	-4.85271×10^{-1}	(ISEN-.2539682540)	3.6
4	-2,56663	(IBOM-.2857142857)(ISEN-.2539682540)	37.2
5	-1.66262	(IBOM-.2857142857)(ICOM-.206349206)	12.2
6	-7.67253×10^{-1}	(ICAR-.2698412698)(ICOM-.206349206)	3.2
7	1.27356×10^{-2}	FPS	6.8
8	2.25967×10^{-2}	(FAN-63.349)	36.0
9	-7.42999×10^{-3}	(FSS-61.138)	9.0
10	2.38503	(UF-1.639	27.0
11	-9.20384×10^{-11}	$(UP-133606.3)^2$	25.0
12	-1.52864×10^{-4}	$(W-64.314)^2$	8.4
13	-1.07105×10^{-3}	$(FAN-48.895)^2$	33.6
14	1.20418×10^{-3}	$(FEM-46.991)^2$	33.6
15	7.10025×10^{-4}	$(FXR-40.172)^2$	10.9
16	-1.61651×10^{-4}	$(FSS-51.898)^2$	2.2
17	-1.11568×10^{-6}	$(PD-722.249)^2$	7.3
18	5.009996	$(UF-1.681)^2$	42.2
19	1.70042×10^{-3}	$(BF-27.288)^2$	13.0
20	4.60293×10^{-1}	ln(UP)	31.4
21	2.35583×10^{-1}	ln(V)	4.8

TABLE V

Westinghouse Model - Phase II Data

$$\ln(LSC/OH) = b_o + \sum_{i=1}^{18} b_i X_i$$

R^2 = .8827 F-Value = 41.0

i	b_i	X_i	Partial-F
0	-6.97950		
1	7.85143×10^{-1}	IFGCOM	10.24
2	1.14876	IBMNAV	34.81
3	1.07719	IBMCOM	21.16
4	1.91500×10^{-1}	CD	12.25
5	-1.22007×10^{-2}	FDI	37.21
6	-1.72307×10^{-2}	FEM	24.01
7	-9.49029×10^{-3}	FXR	4.84
8	-8.36154×10^{-3}	FSS	9.61
9	-3.35635×10^{-4}	(V-1333.0)	9.00
10	1.98641×10^{-2}	(W-32.3)	17.64
11	6.72953×10^{-8}	$(V-3222.0)^2$	6.25
12	-1.05350×10^{-4}	$(W-65.3)^2$	4.00
13	-4.24991×10^{-8}	$(CC-2986)^2$	5.76
14	-4.36525×10^{-4}	$(FPS-45.48)^2$	9.61
15	7.79704×10^{-1}	$(UF-1.72)^2$	16.81
16	5.64131×10^{-1}	ln(UP)	94.09
17	4.61602×10^{-1}	ln(V)	8.41
18	1.47264×10^{-1}	ln(PD)	6.25

TABLE VI

Pulcher's SPSS Model - Phase I Data

$R^2 = 0.95212$ $\bar{R}^2 = 0.92388$ $F = 33.72$

$$\ln (LSC/OH) = \alpha_o + \sum_i \alpha_i D_i + \sum_j \beta_{jo} \ln x_j + \sum_j \sum_i \beta_{ji} D_i \ln x_j$$

Variable No.	Coefficient	Partial F
1	0.402702	13.63
3	0.084548	0.10
5	0.412407	37.28
8	11.320694	23.80
10	-1.135445	17.68
11	-1.457859	26.48
14	3.710527	7.25
16	-2.950970	9.44
17	-0.092716	0.09
20	0.322015	0.07
23	-0.568085	27.14
26	-0.729848	7.51
27	-1.803242	9.46
28	2.506829	12.27
63	-1.995969	18.20
64	3.034970	17.51
68	-0.272142	7.44
70	-0.758240	8.11
75	0.294839	25.70
90	-0.456146	24.86
94	0.697895	25.90
96	-0.642736	43.88
Constant	-5.315378	79.01

TABLE VII

Pulcher's Leaps and Bounds Models - Phase I Data

$$\ln (LSC/OH) = \alpha_0 + \sum_i \alpha_i D_i + \sum_j \beta_{j0} \ln x_j + \sum_j \sum_i \beta_{ji} D_i \ln x_j$$

Variable	C_p Criterion R^2 = 0.9135 $\bar{R}^2$ = 0.88347 F = 31.21		$\bar{R}^2$ Criterion R^2 = 0.9323 $\bar{R}^2$ = 0.9001 F = 29.25	
	Coefficient	Partial-F	Coefficient	Partial-F
UP	0.245908	8.78	0.313871	14.52
W	0.384075	7.75	0.350494	6.86
SF	-1.061926	12.78	-2.878942	14.29
SB	-1.822390	30.26	-2.195891	39.06
DIG			4.381530	4.88
NF*W	-0.431742	31.61	-0.343076	2.10
NF*CC	-0.466254	13.70	-0.470354	15.84
NF*PD	0.738901	16.62	0.672722	14.59
NC*UP	0.285409	5.13	0.254284	4.04
NC*V	-0.334677	4.93	-0.292486	3.92
SF*CC			0.293229	6.30
DIG*UP	-0.584870	12.86	-0.950128	11.70
DIG*V			-0.971576	2.25
DIG*W			2.676919	4.93
DIG*PD	1.081951	15.97	0.553008	2.59
AN*W	0.309271	16.60	0.239272	9.98
EM*W	0.698175	13.89	0.705835	13.47
EM*PD	-0.555855	21.58	-0.545678	20.61
BF*W	0.866668	28.67	0.828916	27.04
BF*%SS	-0.701034	37.03	-0.706378	38.19
Constant	-3.855040	53.44	-4.091618	64.16

All other coefficients are zero.

which prescreens the variables and eliminates those which are unimportant. The Chow Test also determines which subpopulation really had different coefficients. Sixty variables remained and were used in conjunction with the three models.

A stepwise regression procedure using SPSS was used to develop the first model and the Leaps and Bounds Algorithm was used to create the second and third models, the second using $\bar{R}^2$ as a criterion for selection and the third using Mallows' C_p -statistic as a criterion for selection.

All three of these models did a very good job of predicting the old data as determined by the R^2 value, however, in his final conclusion, prediction intervals were computed using the Omnitab computer package [20], and it was determined that both the Leaps and Bounds C_p and the Leaps and Bounds $\bar{R}^2$ model did a better job of prediction than the SPSS model.

Automatic Interaction Detection

It has been suggested that another method of prescreening variables prior to regression is the Automatic Interaction Detection (AID) computer package developed at the University of Michigan's Institute for Social Research and documented by Sonquist and Morgan [33,34]. This technique is primarily used in constructing models on sociological or categorical data and involves a single interval scaled criterion variable and a mixture of interval, ordinal, and nominally scaled predictor variables.

A typical problem in regression analysis is that one cannot always know in advance which transformations such as X_i^2 or $\ln(X_i)$, or interaction terms such as X_iX_j to introduce in the model so that the predictive power of the model is maximized. A larger error term reported in much

of today's research may be partly due to the way in which these predictor variables are combined in the model, and it is this problem of locating specific interaction effects between variables, if in fact they do exist, that is the basis for this investigation. Since AID also determines the variables most important to the model, its main purpose in this investigation will be as a screening device to locate those variables most important to the regression model, thus reducing the number of possible variables considerably.

AID Algorithm and Objective

The AID analysis is somewhat of a branch and bound procedure using analysis of variance technique that is useful in studying the interrelationships among a set of variables and useful in maximizing the predictive power of a multiple regression model. Unlike most multiple regression procedures, linearity and additivity assumptions are not necessary requirements in the AID analysis.

The AID algorithm accomplishes a sequential division of the entire data into subsets based on that split which causes the greatest reduction in the unexplained variability of the criterion variable. On the first iteration, the entire data base is split into two groups around that variable which allows for the minimum within-group variability measured by the sum of squared deviations of the criterion variable from the group means. On each successive iteration, one of the existing groups is split in the same manner as in the first step. This process continues until one of the stopping criteria has been satisfied.

The AID model can be written as:

$$Y_{mi} = \mu_i + \varepsilon_{mi} \qquad \begin{array}{l} m = 1,2,\ldots,n \\ i = 1,2,\ldots,g \end{array} \tag{34}$$

where: Y_{mi} is the m^{th} criterion variable observation in group i

μ_i is the i^{th} group mean

ε_{mi} is the random error of the m^{th} criterion variable observation in group i

This random error term has the same assumptions as the random error term ε_i which was discussed in Chapter II.

An estimate for μ_i is $\bar{Y}_i$, the sample mean of the observations in group i. Letting $\bar{Y}$ be the sample mean for the criterion variable, the total variability in the criterion variable (in AID notation) can be stated as follows:

$$TSST = \sum_{i=1}^{g} \sum_{m=1}^{n_i} (Y_{mi} - \bar{y})^2 \tag{35}$$

This value will be constant for any given set of n observations. Equation (35) can be expanded to:

$$\sum_{i=1}^{g} \sum_{m=1}^{n} (Y_{mi} - \bar{y}) = \sum_{i=1}^{g} \sum_{m=1}^{n} (Y_{mi} - \bar{Y}_i)^2 + \sum_{i=1}^{g} \sum_{m=1}^{n_i} (\bar{Y}_i - y)^2 \tag{36}$$

or: TSST = WSS + BSS

where: TSST is the total sum-of-squares for the entire sample

WSS is the within-group sum-of-squares

BSS is the between-group sum-of-squares

The last term can be simplified to:

$$BSS = \sum_{i=1}^{g} n_i (\bar{Y}_i - \bar{y})^2 \tag{37}$$

The objective of the AID algorithm at each iterative step is to split the groups so that BSS is as large as possible thus making WSS

as small as possible. A good measure of the goodness of the resulting model is:

$$R^2 = \text{BSS*} / \text{TSST} \qquad 0 \leq R^2 \leq 1 \qquad (38)$$

where BSS* is the BSS of the existing groups. As in the multiple regression case, the R^2 value indicates the fraction of the variability in the criterion variable explained by the regression equation. In AID, an R^2 value close to one indicates that the splitting process has done a good job of grouping observations with nearly identical values of the criterion variable.

At each split, equation (34) can be written as:

$$TSS_i = WSS_i + BSS_i \qquad (39)$$

Using this notation, the AID algorithm at each iteration can be generalized as follows:

(1) Select that unsplit sample group which has the largest total sum-of-squares around its own mean as a candidate for further splitting.

(2) For each predictor variable, find the subset of observations in the group selected in Step 1 which maximizes BSS_i (or minimizes WSS_i).

(3) Chose the best partition of observations on a predictor and split the group using that predictor variable.

(4) Repeat Step 1 until a stopping criteria has been satisfied.

The logic of the AID algorithm can be easily summarized in a flow diagram developed by Gooch [14] and simplified by McNichols [25] in Figure 2.

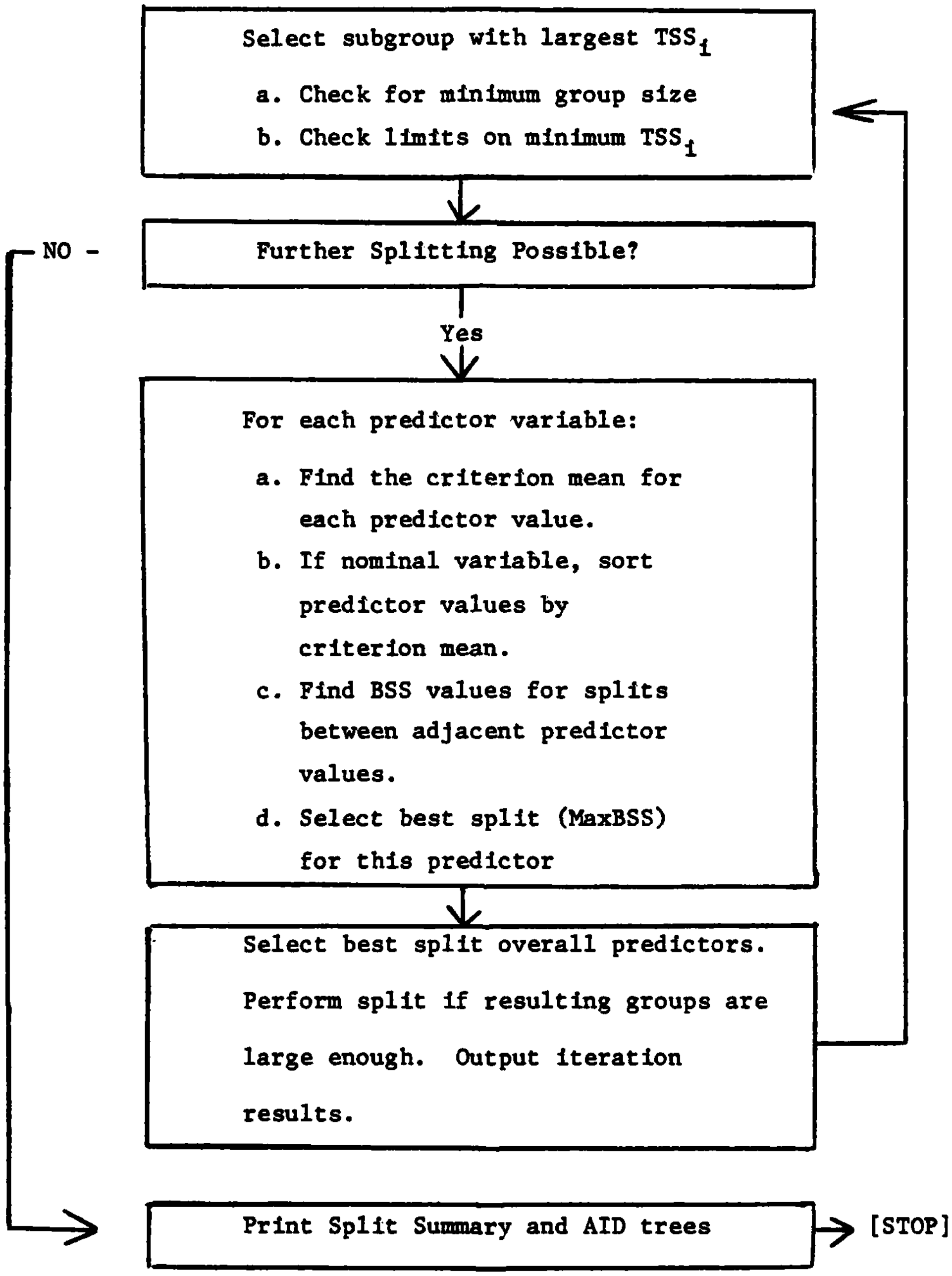


FIGURE 2 Logic of the AID Algorithm

Stopping Criteria

There are four important stopping criteria used in the AID algorithm which are indicated by the user.

(1) The maximum number of final groups including those which can and cannot be further split cannot exceed the value MAXGP or termination will occur.

(2) The number of observations in each group that is split cannot be less than the value NMIN.

(3) The total sum of the squares in a group, TSS_i, cannot be less than P1 percent of the total sum of squares for the entire sample, TSST. Numerically speaking, $P1 < TSS_i/TSST$.

(4) Any split must reduce the original within group sum of squares by P2 percent or the AID algorithm is terminated.

Gooch suggests that:

$P1 \geq .01$

$P2 \geq .005$

$MAXGP \leq 90$

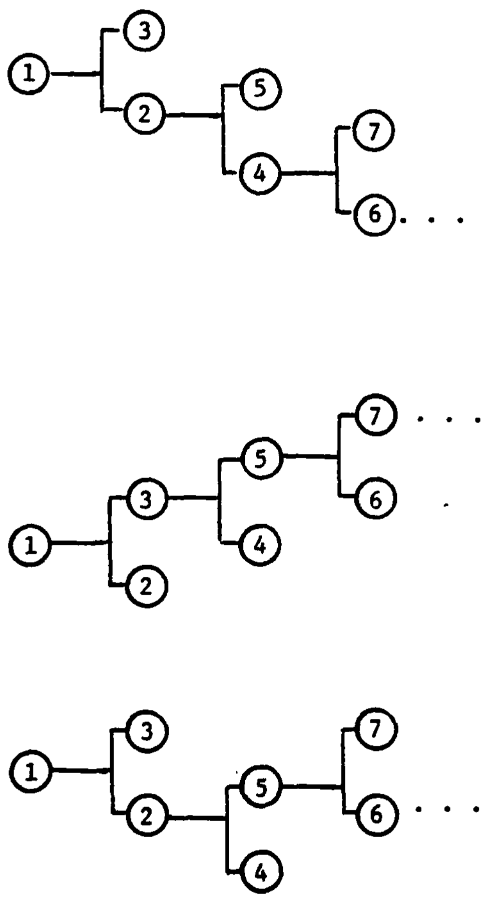
$NMIN \geq 5\%$ of the total number of observations

Analysis of the AID Output

One of the main features of the AID package is the three diagram which graphically describes the splitting process of each of the groups. The structure of these trees is very important in determining the nature of the variable interactions in the model.

Sonquist and Morgan describe two basic structures or shapes of the trees, the trunk-twig structure, and the trunk-branch structure. The truck-twig structure allows only one of two groups split to be split again. The group that is not split is classified a final group.

There are three basic types of trunk-twig structured trees: top termination, bottom termination, and alternating termination (See Figure 3). The top termination structure is referred to by Sonquist as an "alternative advantage" model, where the nature of the advantage is determined by the characteristic which split the group. In this structure, those groups in the upper branches always have a higher mean value than the lower branches, and once formed, these upper branches cannot be split any further.



a. TOP TERMINATION

b. BOTTOM TERMINATION

c. ALTERNATING TERMINATION

FIGURE 3 Trunk-Twig Structured AID Trees

Sonquist refers to the bottom-termination structured tree as an "alternative disadvantage" tree, where the nature of the disadvantage is determined by the characteristic which split the group. In this case, the lower branches once formed, cannot be split further.

In the alternating termination structure, the interpretation can be viewed as a combination of the two preceding structures whereby the importance of a split depends solely on the characteristics of the variable which split the group.

The trunk-branch structure is analogous to the trunk-twig structure except that each group split is a candidate for further splitting. This type of tree structure is typical of the first few splits in any AID tree. Once the first few splits on a group have been made, the structure usually exemplifies that of the trunk-twig structure.

Besides the structure of the tree, the symmetry of the tree, or lack thereof, concerning the extent to which the same variables appear in a split on various trunks is important also. Non-symmetry implies that an interaction exists. Also, if a variable is split on one trunk and shows no indication of reducing the predictive power in another trunk, then there is a clear evidence of an interaction effect between that variable and those used in the preceding splits. The predictive power of each variable in a group is evaluated by the statistic BSS_i/TSS_i and is shown on the selected AID output in Appendix D. This statistic represents the proportion of the variation in the group to which the predictor variable is being applied that would be explained if that group were split.

Preparation for the use of AID

In order to use the AID computer package, several important steps had to be followed. First of all, the data had to be transformed so that an integer format could be used to describe each data element in a six place field. Since many variables were calculated to as many as 13 decimal places, those variables had to be multiplied or divided by a specified factor of 10 and then truncated. For example: LSC/OH was multiplied by 10^4 then truncated, so LSC/OH(27) = 26.63122286176 became 266312.

It is possible that by reducing the number of significant places, round off errors and non-comparible values would result.

Secondly, all data points for each variable had to be sequentially ordered and placed into groups or categories of equal size. (See Table VIII) This is done so that when the groups are split by AID, each mean will be stable with respect to the elements in that group.

After the data is transformed to the proper form, the computer deck can be formed. The itemized input is described in Appendix C.

Results

As stated earlier, the important parameters in the AID input are P1, P2, NMIN, and MAXGP. Many attempts with various combinations of these parameters were made and are described in Table IX.

In the first four runs NMIN was set to 4, which means that no groups will be split unless there are at least 8 data points in that group (4 for each subgroup split).

TABLE VIII

Sequential Ordering of Variables

Variable	No.	Recode	
FGTNAV	1	0	LESS THAN 1
		1	1 OR OVER
BOMNAV	2	0	LESS THAN 1
		1	1 OR OVER
CARNAV	3	0	LESS THAN 1
		1	1 OR OVER
FGTSEN	4	0	LESS THAN 1
		1	1 OR OVER
BOMSEN	5	0	LESS THAN 1
		1	1 OR OVER
FGTCOM	6	0	LESS THAN 1
		1	1 OR OVER
BOBCOM	7	0	LESS THAN 1
		1	1 OR OVER
UNIT PRICE	8	0	LT. OR EQ.TO 2241
		1	2242 TO 3914
		2	3915 TO 8410
		3	8411 TO 19274
			19275 OR OVER
VOLUME	9	0	LT. OR EQ. TO 275
		1	276 TO 560
		2	561 TO 1377
		3	1378 TO 1734
		4	1735 OR OVER
WEIGHT	10	0	LT. OR EQ. TO 850
		1	851 TO 1500
		2	1501 TO 3600
		3	3601 TO 4900
		4	4901 OR OVER
COMPONENTCOUNT	11	0	LT. OR EQ. TO 88
		1	89 TO 399
		2	400 TO 911
		3	912 TO 1186
		4	1187 OR OVER

TABLE VIII (Cont'd)

Variable	No.	Recode
PERCENTDIGITAL	12	0 LT. OR EQ. TO 50
		1 51 TO 440
		2 441 TO 550
		3 551 TO 870
		4 871 OR OVER
PERCENTANALOG	13	0 LT. OR EQ. TO 240
		1 241 TO 740
		2 741 TO 750
		3 751 TO 990
		4 991 OR OVER
PERCENTEM	14	0 LT. OR EQ. TO 5
		1 6 TO 20
		2 21 TO 140
		3 141 TO 760
		4 761 OR OVER
PERCENTPS	15	0 LT. OR EQ. TO 5
		1 6 TO 80
		2 81 OR OVER
PERCENTXMTR	16	0 LT. OR EQ. TO 100
		1 101 TO 190
		2 191 TO 250
		3 251 OR OVER
PERCENTSS	17	0 LT. OR EQ. TO 230
		1 231 TO 860
		2 861 TO 975
		3 976 TO 995
		4 996 OR OVER
POWERDIS	18	0 LT. OR EQ. TO 60
		1 61 TO 150
		2 151 TO 270
		3 271 TO 500
		4 501 OR OVER
BITFIT	19	0 LT. OR EQ. TO 5
		1 6 TO 40
		2 41 TO 130
		3 131 OR OVER

TABLE VIII (Cont'd)

Variable	No.	Recode
IC	20	0 LT. OR EQ. TO 1 1 2 TO 5 2 6 TO 77 3 78 OR OVER
SRU	21	0 LT. OR EQ. TO 3 1 4 TO 9 2 10 TO 12 3 13 TO 16 4 17 OR OVER
QPA	22	0 LESS THAN 2 1 2 OR OVER

TABLE IX

Result of AID Runs

	Run Number							
Parameters	1	2	3	4	5	6	7	8
P1	.015	.01	.0015	.001	.005	.015	.01	.005
P2	.015	.01	.0015	.001	.005	.005	.005	.005
NMIN	4	4	4	4	3	3	3	5
MAXGP	30	30	30	30	30	30	30	30
R^2	.617	.617	.683	.683	.694	.689	.694	.596
Variables*								
V	X	X	X	X	X	X	X	X
AN	X	X	X	X	X	X	X	X
W	X	X	X	X	X	X	X	X
CC	X	X	X	X	X		X	X
CARNAV	X	X	X	X	X	X	X	
XMTR		X	X	X	X	X	X	
PD			X	X	X	X	X	X
UP								X

* Those which AID determined.

This value was lowered to 3 in the following 3 runs. Notice that when NMIN was increased to 5 in run number 8, R^2 decreased from .694 in run number 8 to .596. So indeed, these parameters are important in modeling decisions.

The two best runs (based on highest R^2 values) were runs 5 and 7, where number 7 contains three parameters recommended by Gooch. Run number 7 was chosen as the test case to build the regression model used in this research and two approaches were developed from this run. The AID tree and results for run number 7 are described in Figure 4 and Table X.

Since the main objective of using AID is to reduce the total number of variables used and only choose those which are most important to the regression, a choice can be made as to where to stop considering variables for analysis purposes.

If the analysis is stopped when N reaches 4, then three variables remain: V, W, and AN. Considering interaction terms or cross produce terms, six variables can be used: V, W, AN, V·W, V·AN, and W·AN.

Another choice would be to stop considering variables for analysis when N reaches 3. In this case, 7 variables remain, V, W, CC, PD, AN, XMTR and CARNAV. AN and XMTR can be considered partial indicators in the sense that they can be represented as indicators (0 or 1) where zero indicates that AN or XMTR equals zero and the value one indicates that AN or XMTR is greater than zero. These indicator variables are referred to in the analysis as IAN and IXMTR. CARNAV is a pure indicator (either 0 or 1). In this case it was decided to use interaction terms between the first six original variables and the

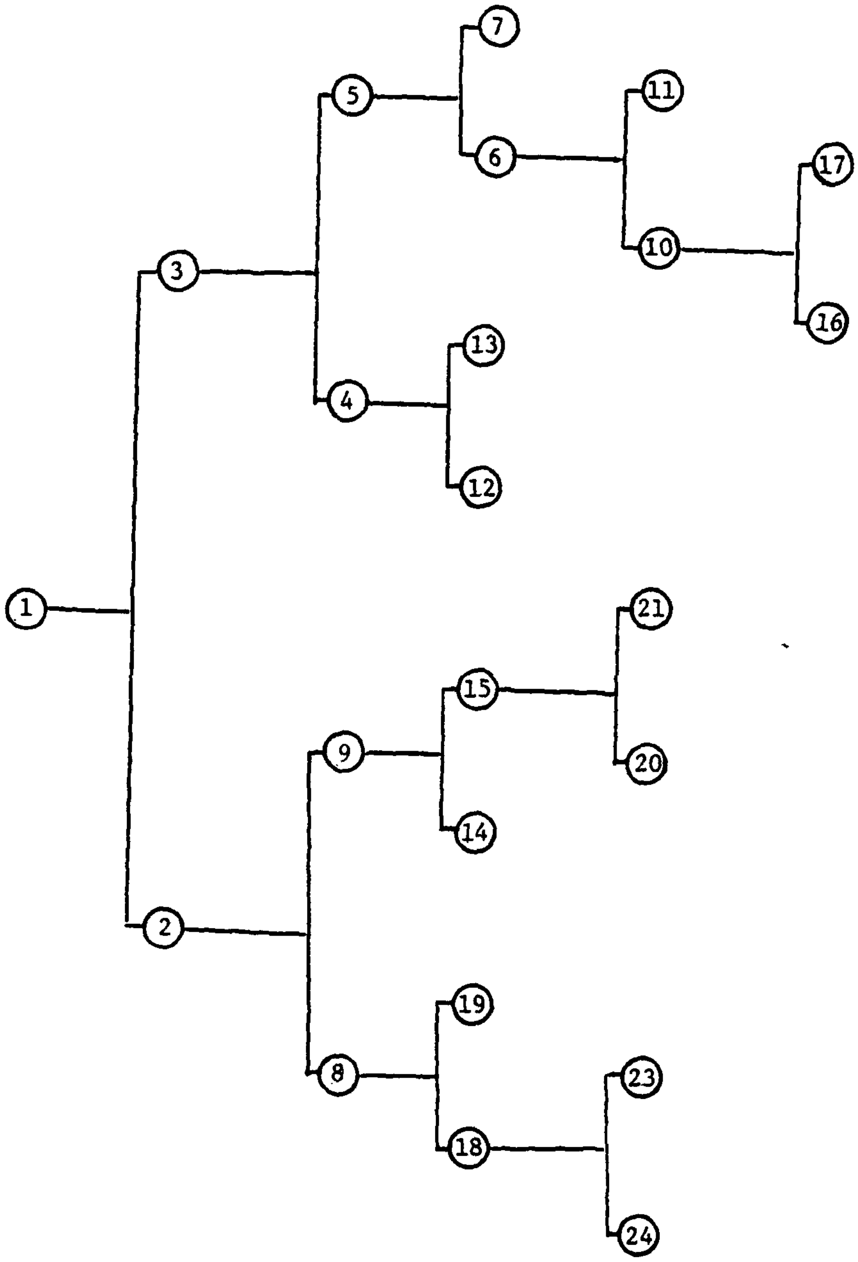


Figure 4 AID Tree for Run No. 7

TABLE X

AID Tree Results for Run No. 7

GROUP	VARIABLE	RECODE	MEAN	STD. DEV.	N	R^2
1	-	-	13687.90	15871.88	63	-
2	V	0 1 2	6708.97	7352.77	38	.294
3	V	3 4	24295.88	19133.53	25	.294
4	AN	1 3 4	11667.33	9052.37	9	.435
5	AN	0 2	31399.44	19640.68	16	.435
6	CC	1 3 4	25506.08	9193.78	12	.540
7	CC	0 2	49079.50	29540.98	4	.540
8	W	0 1 3 4	4040.72	4309.02	29	.595
9	W	2	15306.67	8460.32	9	.595
10	CARNAV	0	21670.25	9449.06	8	.617
11	CARNAV	1	33177.75	4745.71	4	.617
12	CC	2 4	6328.60	3030.33	5	.638
13	CC	1 3	18340.75	9629.98	4	.638
14	AN	3 4	6313.67	1385.32	3	.661
15	AN	0 1 2	19803.17	6763.90	6	.661
16	PD	4	17311.25	7900.37	4	.670
17	PD	2 3	26029.25	6508.13	4	.670
18	XMTR	0 1	2957.56	3052.07	25	.684
19	XMTR	2 3	10810.50	4820.07	4	.684
20	PD	1 4	16137.67	4832.67	3	.689
21	PD	2	23468.67	6417.00	3	.689
22	W	1 3	943.50	670.84	12	.694
23	W	0	4816.69	3208.98	13	.694

three indicators variables. A total of twenty-three variables are created in this case. A list of both sets of variables created in this case are listed in Table XI.

In order to decide which model should be used, each set of variables was run through the IMSL-RLEAP (Leaps and Bounds) program described earlier. Using $\bar{R}^2$ as a criterion, the 23-variable model explained 71.8 percent of the variance with 17 of the 23 variables, while the 6-variable models only explained 50.1 percent of the variance using all six variables. See Appendix D for a selected AID output and Appendix F for a selected Leaps and Bounds output.

Next a log transformation was made on the 23-variable model and run through Leaps and Bounds, and, surprisingly, the results did not show an improvement over those of the untransformed data. Thus, the untransformed 17 variables chosen by Leaps and Bounds were accepted as those AID determined most important. This model will therefore be used in the cross-validation experiments to follow. This 17-variable model is described in Table XII.

TABLE XI

Variables in the Two AID Models Considered

Model 1	Model 2
V	V
AN	W
W	CC
V·W	PD
V·AN	AN
W·AN	XMTR
	CARNAV
	V·CARNAV
	W·CARNAV
	CC·CARNAV
	PD·CARNAV
	XMTR·CARNAV
	V·IAN
	V·IXMTR
	W·IAN
	W·IXMTR
	CC·IAN
	CC·IXMTR
	PD·IAN
	PD·IXMTR
	AN·IXMTR
	XMTR·IAN
	IAN
	IXMTR

TABLE XII

AID Regression Model Determined by Leaps and Bounds

$$LSC/OH = B_o + \sum_{i=1}^{17} B_i X_i$$

$$R^2 = .718$$

i	B_i	X_i	Partial-F
0	2.658290567	-	-
1	$-.155899 \times 10^{-2}$	V	11.7804
2	$-.779107 \times 10^{-1}$	W	22.7635
3	$.105464 \times 10^{-2}$	PD	5.52895
4	$.961796 \times 10^{-1}$	XMTR	8.75757
5	$.261128 \times 10^{1}$	CARNAV	7.58031
6	$.700891 \times 10^{-1}$	W.CARNAV	14.4932
7	$-.506175 \times 10^{-2}$	PD.CARNAV	18.0098
8	$-.267022 \times 10^{-1}$	AN·CARNAV	7.62132
9	$.878194 \times 10^{-3}$	V.IAN	9.5718
10	$-.12007 \times 10^{-2}$	W·IAN	17.8296
11	$.143445 \times 10^{-2}$	W·IXMTR	3.85888
12	$.204243 \times 10^{-2}$	CC.IAN	12.2973
13	$-.112446$	CC.IXMTR	13.4675
14	$.21432 \times 10^{-2}$	PD.IAN	15.1695
15	$-.402166 \times 10^{-2}$	PD.IXMTR	22.0968
16	$.153848$	XMTR.IAN	8.40514
17	$-.547619 \times 10^{1}$	IXMTR	7.77109

V Cross Validation, Conclusions and Recommendations

Cross Validation

Three equations developed by Pulcher and one developed by Westinghouse have been reviewed, and one model developed by AID has been analyzed. All have been based on the old Westinghouse data collected in Phase I containing 63 data points.

A cross validation procedure was used to determine how well these old models predict the new 71 data points contained in the Phase II.

The first step was to use the new data in each of the old models to find the cross validation SSE and SST. They were then used to find the cross validation R^2 described in Chapter II. A summary of results is given in Table XIII.

In both the Westinghouse model and the AID model, the cross validation SSE was greater than the SST. This would tend to imply that neither of the two models predict the new data very well. This is a surprising result especially for the Westinghouse model.

One possible explanation for this is that the Westinghouse model was developed in such a way that much of the idiosyncrecies of the data were explained. Notice the vast difference between the first model described in Table IV and the second described in Table V. This could also be the reason why the AID model failed to predict the new data.

TABLE XIII

Cross Validation Results

Model	c.v. SSE	SST	c.v. R^2
L & B - $\bar{R}^2$	157.9096802275	227.701363	.3065053361
L & B - C_p	112.4418454273	227.701363	.50618720
SPSS	89.457779054	227.701363	.6071269467
Westinghouse	*	227.701363	*
AID	*	1755.2523798	*

* c.v.SSE was greater than SST

The best model determined by the cross validation criterion was Pulcher's SPSS model which had a c.v. R^2 value of .607 (see Table XIII). Using those variables, updated coefficients have been computed (see Table XIV). This new model using the old variables and just the Phase II data has an R^2 value of .780 indicating that 78% of the variance in the dependent variable is explained by the model. With the complete set of data (134 data points) 70.9% of the variance was explained by the model. Table XV describes this model. (See Appendix E for selected outputs from SPSS.)

Conclusions

A review of past research indicates that much literature is available on criterion in the selection of variables in a multiple regression thus indicating that it is an important subject not only for mathematicians or operations researchers, but is important to anyone attempting to develop valid models both for description and prediction purposes. As a result, these criteria give the statisticians a useful index of how well various models fit the data, however, experience shows that the result of using a single criterion should not be accepted as a final answer, but should be used with other available statistics and individual's intuitive judgement in developing a sound analysis.

This cross validation R^2 value was useful in evaluating the prediction capabilities of the five models discussed. The three models which used log transformed data and were developed by Pulcher for description purposes on the old Westinghouse data

TABLE XIV

Pulcher's SPSS Model fitted to the New Data Points

$$\ln LSC/OH = \alpha_o + \sum_i \alpha_i D_i + \sum_j \beta_{jo} \ln x_j + \sum_j \sum_i \beta_{ji} D_i \ln x_j$$

$\bar{R}^2$ = .67923 R^2 = .78004 F = 7.73752

	Variable Name	Coefficient	Partial-F
	UP	.36000615	.17946696
	W	.60315963	.43885436
	SS	1102.0708	280.30031
	NB	8.2056618	17.346404
	SF	-.33287310	.39887747
	SB	.99001459	.83450842
	DIG	-1.3736140	2.5219780
	EM	2.6000225	1.6961873
	PS	.12680075	.31751393
	NF * UP	.13008302	.15751183
	NF * CC	-.13099197	.22804494
	NB * UP	-.34773058	1.1390687
**	NB * V	-	-
	NB * W	-.75005236	2.6271934
	DIG * V	-.14280299	.60250900
	DIG * W	.49276704	.61187387
	AN * UP	.06467638	.13817898
	AN * W	-.13646127	.43869986
	EM * V	-.35382193	.25552978
	XMTR * CC	-.36370529	.46191312
	XMTR * SS	531.35247	667.96390
	BF * W	.0559571	.2229430
	BF * SS	59.533507	207.10304
	Constant	-10.43849	1.4928598

**Removed from the equation by SPSS.

TABLE XV

Pulcher's SPSS Model Variables fitted to the Entire Data Set

(Phase I & Phase II)

$$\ln LSC/OH = \alpha_o + \sum_i \alpha_i D_i + \sum_j \beta_{jo} \ln x_j + \sum_j \sum_i \beta_{ji} D_i \ln x_j$$

$\bar{R}^2 = .64868$ $R^2 = .70944$ $F = 11.67717$

Variable Name	Coefficient	Partial-F
UP	7.1760834	2.6076596
W	- .58168533	2.8060750
SS	- .31049321	.45804725
NB	- .12129190	.00539864
SF	-1.0966765	.78704103
SB	.20884626	.69860270
DIG	.16839348	3.1550771
EM	.50533178	16.807697
PS	.43488570	2.1115964
NF * UP	.036039607	.13695163
NF * CC	- .048746736	.11610925
NB * UP	- .11413454	.22280423
NB * V	-1.7249398	2.2502326
NB * W	2.17249398	2.2513077
DIG * V	- .19976096	.23782540
DIG * W	.32683368	.50702946
AN * UP	.024885722	.0627402
AN * W	.032724716	.012309140
EM * V	.14565893	.59932259
XMTR * CC	.29069072	7.6961250
XMTR * SS	- .38171889	5.9613262
BF * W	.028545024	.31109000
BF * SS	- .008700618	.0471961
Constant	- .70977903	85.146786

had adequate predictive capabilities; the other two models (the Westinghouse model and the AID model) were determined not to have very good predictive capabilities.

The Automatic Interaction Detection Algorithm was useful in prescreening important variables and reducing the total number of variables to be used in a multiple regression, however, it did not prove to be the best technique in developing regression models, for the maximum R^2 value was only .780.

Recommendations

In his research, Pulcher used the Chow Test as a screening device to determine the most important variable subset using a Product of Powers model. However, one assumption in using the test is that of equal variances on the error term. In future analysis, I would recommend the use of a technique developed by Jayatissa [21] of Tests of Equality Between Subsets of Coefficients in Two Multiple Regressions assuming unequal variances. This can be used as a prescreening device to locate important variables. Then stepwise regression procedures using SPSS can be used to develop a multiple regression model.

To the personnel at the Avionics Laboratory, I would recommend that cross validation studies be made to insure that models developed by contractors be able to predict new data so that new models do not have to be developed every time new data is obtained.

All techniques used on this analysis were based on minimizing the sum of squared errors. The many criterion for selection of variables mentioned in this report should be given further consideration.

Bibliography

1. Aitken, Murray A. "Simultaneous Inference and the Choice of Variable Subsets in Multiple Regression", Technometrics, 16: 221-227 (May 1974).

2. Allen, David M. "Mean Square Error of Prediction as a Criterion for Selecting Variables", Technometrics, 13: 469-475 (August 1971).

3. Allen, David M. "The Relationship Between Variable Section and Data Augmentation and a Method for Prediction", Technometrics, 16: 125-127 (February 1974).

4. Andrews, D.F. "A Robust Method for Multiple Linear Regression", Technometrics, 16: 523-531 (November 1974).

5. Beale, M.M.L. et al. "The Discarding of Variables in Multivariate Analysis", Biometrika, 54: 357-366 (1967).

6. Bendel, Robert B. and A.A. Afifi. "Comparison of Stopping Rules in Forward 'Stepwise' Regressions", Journal of the American Statistical Association, 72: 46-53 (March 1977).

7. Chow, Gregory C. "Tests of Equality Between Sets of Coeffecients in Two Linear Regressions", Econometrica, 28: 591-605 (July 1960).

8. Daniel, Cuthbert and Fred S. Wood. Fitting Equations to Data. New York: John Wiley and Sons, Inc. (1971).

9. Diehr, George and Donald R. Hoflin. "Approximating the Distribution of the Sample R^2 in Best Subset Regressions", Technometrics, 16: 317-320 (May 1974).

10. Draper, N.R. and H. Smith. Applied Regression Analysis. New York: John Wiley and Sons, Inc. (1966).

11. Ellerton, Roger R.W. "Is the Regression Equation Adequate? - A Generalization", Technometrics, 20: 313-315 (1978).

12. Fisher, Franklin M. "Tests of Equality Between Sets of Coefficients in Two Linear Regressions: An expository Note", Econometrica, 38: 361-366 (March 1970).

13. Furnival, George M. and Robert W. Wilson. "Regression by Leaps and Bounds", Technometrics, 16: 499-511 (November 1974).

14. Gooch, L.L. Policy Capturing With Local Models: The Application of the AID Technique in Modeling Judgement, Unpublished Ph.D. Dissertation, University of Texas, Austin, Texas, 1972.

15. Garland, J. "A Relatively Simple Form of the Distribution of the Multiple Correlation Coefiicient". J. Roy. Statis. Soc. B30, 276-283 (1968).

16. Helms, Ronald W. "The Average Estimated Variance Criterion for the Selection-of-Variables Problem in General Linear Models", Technometrics, 16: 261-273 (May 1974).

17. Hocking, R.R. "Criteria for Selection of a Subset Regression: Which One Should be Used", Technometrics, 14: 967-970 (1972).

18. Hocking, R.R., "The Analysis and Selection of Variables in Linear Regression", Biometrics, 32: 1-49 (March 1976).

19. Hocking, R.R., and R.N. Leslie. "Selection of the Best Subset in a Regression Analysis", Technometrics, 9: 531-540 (1967).

20. Hogben, David et al. OMNITAB II Users Reference Manual, Washington, D.C.: National Bureau of Standards (1971).

21. Jayatissa, W.A. "Tests of Equality Between Sets of Coefficients in Two Linear Regressions when Disturbances are Unequal", Econometrica, 45: 1291-1292 (1977).

22. Kennedy, W.J. and T.A. BanCroft. "Model Building for Prediction in Regression Based upon Repeated Significance Tests", Ann. Math. Statist., 42: 1273-1284 (1971).

23. Koplyay, J.B., C.D. Gott, and J.H. Elton, Automatic Interation Detector-Version 4 (AID)-4 Reference Manual, Air Force Human Resources Laboratory, Brooks AFV, Texas, October 1973. (Defense Documentation Center AD-773 803).

24. Mantel, Nathan. "Why Stepdown Procedures in Variable Selection", Technometrics, 12: 621-265 (1970).

25. McNichols, Charles W. "An Introduction to: Apllied Multivariate Data Analysis". Unpublished Course Notes. The Air Force Institute of Technology, Wright Patterson AFB (1978).

26. Mosier, Charles I. "Symposium: The Need and Means of Cross-Validation", Educational and Psychological Measurement, Vol II: 5-11 (1951).

27. Narule, Subhask C. and John F. Wellington. "Selection of Variables in Linear Regression Using the Minimum Sum of Weighted Absolute Errors Criterion", Technometrics, 21: 299-306 (1979).

28. Nie, Normal H. et al. Statistical Package for the Social Sciences, New York: McGraw Hill Book Company (1975).

29. Park, Sung H. "Selecting Contrasts Among Parameters in Scheffe's Mixture Models: Screening Components and Model Reduction", Technometrics, 20: 273-279 (1978).

30. Pulcher, Larry J. "Criterion for Selection of Variables in a Regression Analysis", Unpublished Master's Thesis, The Air Force Institute of Technology, Wright Patterson AFB (1978).

31. Schatzoff, M., R. Tsao, and S. Fienberg. "Efficient Calculation of All Possible Regressions", Technometrics, 10: 769-779 (1968).

32. Scheffe, H. "Experiments with Mixtures", J. Roy. Statist. Soc., B, 20: 344-360 (1958).

33. Sonquist, J.A. Multivariate Model Building, the Validation of a Search Strategy, Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, 1970.

34. Sonquist, J.A. and J.N. Morgan. The Detection of Interaction Effects, Survey Research Center, Nongraph 35, Institute for Social Research, University of Michigan, Ann Arbor, 1964.

35. Theil, H. Principles of Econometrics, John Wiley and Sons, New York (1971).

36. Turek, John P., E. Louis Wienecke, III and Dr. Erasums E. Feltus, Predictive Operations and Maintenance Cost Model for Air Force Avionics Laboratory. Westinghous Electric Corporation (1979).

37. Webster, J.T., R.F. Gunst, and R.L. Mason, "Latent Root Regression Analyisis", Technometrics, 16: 513-522 (November 1974).

APPENDIX A

LRU DESCRIPTION

APPENDIX A

LRU Description

No.	LRU-ID	AIRCRAFT	DESCRIPTION
1	71B2Ø	F4E	Amplifier, Computer
2	7353Ø	F4E	Ballistics, Computer
3	71LBØ	F4E	Receiver-Transmitter
4	71HKØ	F4E	Platform, Gyro, Stab.
5	71PKØ	RF4C	Receiver-Transmitter
6	71PBØ	RF4C	Amplifier, P.S. RCVR
7	7171Ø	RF4C	P.S. Leveling, Amplifier
8	724GØ	RF4C	Power Supply
9	71G5Ø	RF4C	Computer, Navigation
10	71FAØ	F15A	Amplifier, Electronic
11	71FBØ	F15A	Gyroscope, Displacement
12	71CAØ	KC135A	Receiver-Transmitter
13	71DAØ	F15A	Receiver-Transmitter
14	71ABE	B52H	Receiver
15	71ADA	B52H	Receiver-Transmitter
16	73DBA	B52H	Receiver-Transmitter
17	71ACC	B52H	Receiver
18	73CBØ	B52H	Amplifier
19	73CEN	B52H	Computer, A2 and EL
20	73CFK	B52H	Receiver-Transmitter
21	73DAH	B52H	Amplifier, Electronic Control
22	73EBA	B52H	Amp, Astrotrack, Servo
23	73EBF	B52H	Signal Amplifier
*24	71CAØ	F15A	Receiver
25	72EAA	KC135A	Receiver-Transmitter
26	72ECA	KC135A	Amplifier, Electronic Control
27	72BPO	C5A	Measurement Unit, IMU
28	71JAØ	C5A	Receiver, VHF Navigational
29	71LAØ	C5A	Receiver-Transmitter
30	72DNØ	C5A	Processor Data

* DUPLICATE LRU-ID -- Placed on a Different Aircraft

APPENDIX A

LRU Description (Con't)

No.	LRU-ID	AIRCRAFT	DESCRIPTION
31	72ACØ	C5A	P.S., Thermal Control
32	7171A	C130E	Receiver
33	7131D	C130E	Receiver-Transmitter
34	72RFØ	C130E	P.S. Power Supply
35	72RBØ	C130E	Amplifier
36	51EAØ	F15A	Computer, Air Data
37	52AAØ	F15A	Computer, Flight Control
38	52ABØ	F15A	Computer, Flight Control
39	63BDØ	F15A	Control Panel, Int Nav
40	71AEØ	F15A	Inertial Measurement Unit
41	71AKØ	F15A	Control Indicator, Nav
42	74JAØ	F15A	Indicator, Multiple Air Nav
43	74JCØ	F15A	Processor, Signal Data
44	52GA1	F106	Amplifier-Interface
45	71JCE	C5A	Control Panel VHF Nav
46	72AEØ	C5A	Computer-Primary, IDNE
47	72CCØ	C5A	Computer-Analog/Digital
48	71ZAØ	C130E	Receiver-Transmitter
49	71ZBØ	C130E	Digital/Analog Converter
50	71ZDØ	C130E	Control Unit
*51	71ZAØ	F111D	Receiver-Transmitter
*52	71ZBØ	F111D	Digital/Analog Converter
53	71ZCØ	F111D	Control
54	73EGØ	F111D	Computer, General Purpose
55	73EPØ	F111D	Converter-Multiplexer
56	73HAØ	f111D	Stabilizer Platform
57	73HCØ	F111D	Navigational Computer
58	73NAØ	F111D	Indicator, Horizontal Display
59	73NBØ	F111D	Processor, Horizontal Display
60	73QBØ	F111D	Electronic Unit, Radar

* DUPLICATE LRU-ID -- Placed on a Different Aircraft

APPENDIX A

LRU Description (Con't)

No.	LRU-ID	AIRCRAFT	DESCRIPTION
61	73SCO	F111D	Indicator, Digital Display
62	73KBØ	F111D	Antenna-Receiver
63	73KEØ	F111D	Amplifier, Power Supply
64	73KFØ	F111D	Synchronizer-Transmitter
65	73DDØ	F111D	Computer, Terrain Following
*66	71CAØ	FB111A	Receiver Unit
67	73EGØ	FB111A	Computer, General Purpose
68	73HCØ	FB111A	Navigational Computer Unit
69	73LAØ	FB111A	Electronic Unit
70	7593Ø	F4E	Weapons Release Control
71	74BDØ	F4E	Computer
72	74BFØ	F4E	Transmitter
73	7481Ø	F4E	Gyroscope, Lead Comp.
74	76A1Ø	RF4C	Analyzer, Pulse
75	76GAØ	RF4C	Signal Processor
76	74FFØ	F15A	Processor
77	74FAØ	F15A	Transmitter
78	74FHØ	F15A	Power Supply
79	74FUØ	F15A	Antenna
80	77ECØ	B52H	Flir Signal Proc.
81	77EEØ	B52H	Flir Turret Drive
82	77DCA	B52H	STV Camera, Electronic
83	77DBØ	B52H	STV Turret Drive
84	73CRØ	F4E	Laser Control, Electronic
85	73CGØ	F4E	Two Axis Gimbal Assembly
86	65BHØ	F15A	Processor, Radar Target Data
87	74FCØ	F15A	Receiver, Radar
88	74FJØ	F15A	Oscillator-RF
89	74FKØ	F15A	Radar Set Control
90	74FQØ	F15A	Processor, Radar Data

* DUPLICATE LRU-ID -- Placed on a Different Aircraft

APPENDIX A

LRU Description (Con't)

No.	LRU-ID	AIRCRAFT	DESCRIPTION
91	74KAØ	F15A	Display Unit, Head Up
92	74KCØ	F15A	Processor Signal Data
93	75AEØ	F15A	Converter-Programmer
94	74CAØ	F4E	Indicator, Control
95	74CBØ	F4E	Indicator, Pilot
96	74CCØ	F4E	Indicator, PSO, IO
97	74FA1	F106	-
98	74EBØ	F15A	Lead Computing Gyro
99	76AEA	B52H	Transmitter
100	73KAØ	FB111A	Computer, TFR
101	73PHØ	F111D	Power Supply, LV
102	73PBØ	F111D	Processor, Electronic
103	73PDØ	F111D	Radar Transmitter
104	73PFØ	F111D	Signal Data Converter
105	73PMØ	F111D	Reference Signal Gen.
106	71NA0	F4E	Receiver-Transmitter
107	71QUØ	RF4C	Receiver-Transmitter
108	63AAØ	F15A	Receiver-Transmitter
109	65AA0	F15A	Receiver-Transmitter
110	63BAA	B52H	Receiver-Transmitter
111	63CAA	B52H	Receiver-Transmitter
112	65BAA	B52H	Receiver-Transmitter
113	61BBA	B52H	Receiver
114	65BAA	KC135A	Receiver-Transmitter
115	63AF0	KC135A	Receiver-Transmitter
116	63AA0	C5A	Receiver-Transmitter
117	63121	C130E	Receiver-Transmitter
118	63AAA	C130E	Receiver-Transmitter
119	55ALØ	C5A	Central Multiplex Adapter
120	55AVØ	C5A	Computer Digital, Madar

APPENDIX A

LRU Description (Con't)

No.	LRU-ID	AIRCRAFT	DESCRIPTION
121	61AAØ	C5A	Exciter Receiver, HF/SSB
122	c1ACØ	C5A	Amplifier/Antenna Coupler
123	61AEØ	C5A	Panel, Control, HF/SSB
124	62AAØ	C5A	Transceiver, VHF Comm
125	63A6Ø	F15A	Radio Receiver
126	63BCØ	F15A	Control Panel, Int Comm
127	63BFØ	F15A	Control Panel, IFF
*120	61AAØ	FB111A	Receiver-Transmitter
129	61ABØ	FB111A	Amplifier-Power Supply
*130	61ACØ	FB111A	Control
131	72AAØ	FB111A	Control, Radar Transponder
132	72ACØ	FB111A	Receiver Transmitter
133	64211	C130E	Intercom Set
134	64212	C130E	Control Panel

* DUPLICATE LRU-ID -- Placed on a Different Aircraft

APPENDIX B: PART 1

LISTING OF PHASE I DATA

```
N  LRU-ID  BOMBER  CARGO  SENSORY  COMM
       UP   V   W   CC
       DIG   AN   FM   PS   XMTR
SS   PD   BITFIT   1C   SPU   QPA   LSC/OH
UF

 1 71B20       0.       0.       0.       0.
        13721.     443.     17.70     410.
           0.0       86.0       14.0       0.0       0.0
 86.0     200.       0.0       0.       9.       1.     .59344195+7749
 2.30
 2 73530       0.       0.       0.       0.
        41334.     996.     36.60     2176.
           0.0       73.0       27.0       0.0       0.0
 73.0     175.       4.0       3.       21.       1.     .5930947273960
 2.30
 3 71LB0       0.       0.       0.       0.
         6581.     1444.     40.00     491.
           0.0       75.0       0.0       0.0       20.0
 27.0     212.       0.0       0.       16.       1.     .8404635651393
 2.30
 4 71H60       0.       0.       0.       0.
        35313.     1576.     30.00     73.
           0.0       24.0       76.0       0.0       0.0
 24.0     820.       0.0       0.       4.       1.     6.0677759434447
 2.30
 5 71PK0       0.       0.       0.       0.
         8410.     1473.     40.00     653.
           0.0       74.0       1.0       0.0       20.0
 95.8     77.       0.0       0.       13.       1.     1.1157731131456
 2.30
 6 71PB0       0.       0.       0.       0.
        22+1.     1270.     38.50     758.
           0.0       56.0       14.0       0.0       0.0
 68.0     239.       0.0       0.       19.       1.     .4366631115666
 2.30
 7 71710       0.       0.       0.       0.
          840.       91.     4.00     12.
           0.0       100.0       0.0       0.0       0.0
100.0     20.       1.0       0.       6.       1.     .0045139F351492
 2.30
 8 724G0       0.       0.       0.       0.
         2055.     133.     1.25     84.
           0.0       37.0       3.0       0.0       0.0
 97.0     87.       0.0       0.       1.       1.     .6392870334474
 2.30
 9 71G50       0.       0.       0.       0.
        23327.     584.     14.00     421.
           0.0       80.0       11.0       0.0       0.0
 34.0     172.       13.0       0.       21.       1.     .5175632712515
 2.30
10 71FA0       0.       0.       0.       0.
        17232.     142.     13.90     412.
           +9.0       +9.0       2.0       0.0       0.0
 98.0     334.       19.0     255.       16.       1.     .4554556370743
 2.30
```

```
N  LRU-ID  BOMBER  CARGO  SENSORY  COMM
       UP   V   W   CC
       DIG  AN  EM  FS  XMIT
SS  PO  BITFIT  IC  SFU  OFA  LSC/OH
UF

12 710A0      1.      ?.      C.      0.
          2176.    23?.     6.50    150.
               0.0   190.0       0.0     0.C      ?.C
    100.0      20.     13.J      0.0      9.0      1.0    .0788460000000
     1.30
13 710A?      ?.      ?.      ?.      ?.
         301?5.    83?.     29.80     32.
               0.0   7?.0        0.0     0.0     2?.0
     33.0      70.     18.0      0.0     11.?      1.0   1.0343487916725
     2.30
14 71ABE      1.      0.      0.      0.
          1170.    23?.     7.50    21?.
               0.0    93.0       7.?     0.?      ?.0
     93.0       7.      4.0      0.0      1.0      1.0    .1593542778450
     1.30
15 71ADA      1.      0.      0.      0.
          2744.   1732.    60.00     92?.
               0.0   10?.0       0.0     0.?      ?.0
      0.0      17.      0.0      0.0     17.?      1.0   3.5010442505390
     1.30
16 73D9A      1.      0.      0.      0.
          5928.   6478.    90.00    321.
               0.0    75.0       0.0     0.0     2?.0
     25.0    1640.      0.0      0.0     14.?      1.0   3.0183202097964
     1.30
17 71ACC      1.      ?.      0.      0.
           153.     9?.     2.60     78.
               0.0   100.0       0.0     0.0      ?.0
      0.0      17.      0.0      0.0      3.?      1.0    .2278?26256567
     1.30
18 73CBQ      1.      0.      0.      0.
           1?8.     3?.     1.20     33.
               0.0   100.0       0.0     0.0      ?.0
      0.0      34.      4.0      0.0     22.?      1.0    .0433163408965
     1.30
19 73CEN      1.      ?.      0.      0.
          2752.    256.    10.50     20.
               0.0      0.0    100.0     0.0      ?.0
      0.0     100.      0.0      0.0      1.?      1.0    .3541?71728635
     1.30
20 73CFK      1.      0.      0.      0.
         15720.   323?.   110.00    561.
               0.0    75.0       0.0     0.0     2?.0
     66.5     152.      0.0      0.0     13.?      1.0   8.8305512234367
     1.30
21 73DAH      1.      ?.      0.      0.
          5720.   306?.    58.00    156.
               0.0    43.0      57.0     0.?      ?.0
      0.0     225.      0.0      0.0      1.?      1.0   1.3371??2831169
     1.30
```

```
N  LRU-ID  BOMBER  CARGO  SENSORY  COMM
      UF   V   W   CC
      DIG   AM   FM   FS   XMTR
SS  PD  BITFIT  IC  SPU  OPA  LSC/OH
UF

22 73E3A      1.      0.      0.      0.
        1347.     132.     3.40     120.
            0.0    100.0       0.0      0.0      0.0
   85.0      34.      0.0      0.0      4.0      1.0    .1503180063939
   1.30
23 73E8F      1.      0.      0.      0.
        2530.     464.    11.50     177.
            0.0    100.0       0.0      0.0      0.0
   37.0     130.      0.0      0.0      5.0      1.0   1.3327361151447
   1.30
24 71CA0      1.      0.      0.      0.
        3255.    1730.    60.00     924.
            0.0     75.0       0.0      0.0     25.0
    0.0     500.      0.0      5.0     15.0      1.0   3.5556396185122
   1.30
25 72EAA      0.      1.      0.      0.
        3710.    6478.    87.50     321.
            0.0     75.0       0.0      0.0     25.0
   25.0    1640.      0.0      0.0     39.0      1.0   3.1125257254178
   1.20
26 72ECA      0.      1.      0.      0.
        2351.    4283.    63.36    4363.
            0.0    100.0       0.0      0.0      0.0
    0.0     160.      0.0      0.0      1.0      1.0    .4363912059963
   1.20
28 71JA0      0.      1.      0.      0.
        6247.     479.    12.80    1013.
            0.0     99.0       1.0      0.0      0.0
   99.0     175.      0.0      0.0     10.0      2.0    .9300370135178
   1.20
29 71LA0      0.      1.      0.      0.
       34352.    1250.    32.00    2743.
           55.0     33.0       0.0      0.0     12.0
   99.6     265.      0.0    607.0     17.0      2.0   3.1232286512515
   1.20
30 720NC      0.      1.      0.      0.
      100450.    1511.    39.00    1044.
           37.0     12.0       1.0      0.0      0.0
   38.7     850.      0.0      0.0     12.0      2.0   3.9275376976574
   1.20
31 72AC0      1.      0.      0.      1.
       15793.     432.    31.00     552.
            0.0     99.0       1.0      0.0      0.0
   99.0     851.      4.0     11.0     13.0      1.0    .9183568447272
   1.30
32 71718      0.      1.      0.      0.
        1258.     230.     7.50     214.
            0.0    100.0       0.0      0.0      0.0
  100.0      20.      4.0      0.0      1.0      1.0    .1495334011632
   1.20
```

```
 N  LRU-ID  BOMBER  CARGO  SENSORY  COMM
        UP   V   W   CC
        DIG   AN   FM   PS   XMTR
SS   PD  BITFIT   IC   SRU   OPA   LSC/OH
UF

 33 71310       0.       1.       0.      0.
          2745.     1734.     60.00      924.
               0.0      71.0          0.0        0.0       27.0
     0.0       500.         0.0        0.0       35.0       1.0    2.6654579275626
    1.20
 34 72RF0       0.       1.       0.      0.
          1532.     339.     14.00      323.
               0.0      0.0          0.0      110.0        0.0
    30.0       860.         0.0        0.0        4.0       1.0     .3789121561943
    1.20
 35 72R90       0.       1.       0.      0.
          2135.     350.      5.00       83.
               0.0      33.0        67.0       0.0        0.0
     0.0       860.         0.0        0.0        1.0       1.0     .3765543312362
    1.20
 66 710A0       1.       0.       0.      0.
          2079.     270.      7.30      290.
               0.0     100.0         0.0       0.0        0.0
   100.0         6.        0.0        0.0       14.0       1.0     .1267372536674
    1.30
 70 75930       0.       0.       1.      0.
          3015.      34.      4.00      103.
               0.0     100.0         0.0       0.0        0.0
   100.0         7.        0.0        0.0        2.0       1.0     .3947257336033
    2.30
 71 74800       0.       0.       1.      0.
         10120.    1509.     43.70     1125.
               0.0      51.0        39.0       0.0        0.0
    61.0       212.         0.0        2.0       17.0       1.0    1.2233016234990
    2.30
 72 74BF0       0.       0.       1.      0.
         15258.    1377.     78.50      393.
               0.0       0.0         0.0       0.0       10.0
    38.0       590.         1.0        5.0        5.0       1.0     .7510315714622
    2.30
 73 74810       0.       0.       1.      0.
          6257.     577.     11.00       35.
               0.0       0.0       100.0       0.0        0.0
     0.0       430.         5.0        0.0        1.0       1.0     .2109392357225
    2.30
 74 76A10       0.       0.       1.      0.
          2552.     550.     15.00      911.
               0.0     130.0         0.0       0.0        0.0
   100.0       335.-        0.0        1.0       10.0       1.0     .3583121228120
    2.30
 75 76G40       0.       0.       1.      0.
         19274.     331.     25.00     1313.
              100.0       0.0         0.0       0.0        0.0
   100.0      1300.         0.0      469.0       11.0       1.0    2.2537575721467
     .30
```

```
N  LRU-ID  BOMBER  CARGO  SENSORY  COMM
      UP  V  W  CC
      DIG  AN  FM  FS  XMTR
SS  PD  BITFIT  IC  SPU  QPA  LSC/OH
UF

76 74FF0       0.      0.      1.      0.
      220943.    2278.    41.00    7631.
        130.0      0.0      0.0      0.0      0.0
   100.0    1300.    51.0   4625.0    37.0    1.0    .8368324407471
    2.30
77 74FA0       0.      0.      1.      0.
      156330.    5330.   173.70    1553.
         0.0      0.0      0.0      0.0    100.0
    99.0     270.    61.0    100.0    10.0    1.0   2.6223657407070
    2.30
78 74FH0       0.      0.      1.      0.
      42023.    1366.    35.00     932.
         0.0      0.0      0.0    110.0      0.0
   100.0    1620.    65.0     13.0     5.0    1.0   1.3431374566476
    2.30
79 74FU0       0.      0.      1.      0.
      142554.    3556.   110.00     13.
         0.0      0.0    100.0      0.0      0.0
     0.0     400.    40.0      0.0    14.0    1.0   3.8331214849123
    2.30
80 77EC0       1.      0.      1.      0.
       9721.    1760.    45.00     945.
        11.0     34.0      5.0      0.0      0.0
    94.7     350.     0.0     37.0    13.0    1.0   1.3352117932307
    1.30
82 77DCA       1.      0.      1.      0.
      31598.    1223.    20.40    1457.
         0.0    100.0      0.0      0.0      0.0
   100.0     145.     0.0      5.0    13.0    1.0    .4339374066279
    1.30
83 77D90       1.      0.      1.      0.
        335.     222.     8.50      0.
         0.0      0.0    100.0      0.0      0.0
     0.0     100.     0.0      0.0     3.0    1.0    .0139233405203
    1.30
84 73CR0       0.      0.      1.      0.
      24542.     337.     8.00     523.
         0.0     96.0      2.0      0.0      0.0
    98.0     300.     0.0     89.0    13.0    1.0    .0574312373524
     .83
85 73CG0       0.      0.      1.      0.
      43912.     970.    12.00      63.
         0.0      0.0    100.0      0.0      0.0
     0.0      60.     0.0      0.0     2.0    1.0    .1559286550591
     .83
106 71NA0      0.      0.      0.      1.
       7131.    1317.    36.00    1074.
         0.0     75.0      0.0      0.0     20.0
    97.6     256.     0.0      0.0    11.0    1.0   2.3557301827467
    2.30
```

```
N  LRU-ID  BOMBER  CARGO  SENSORY  COMM
       UP   V   W   CC
       DIG   AN   EM   PS   XMTR
SS   PD  BITFIT  IC   SRU  QPA   LSC/OH

107 710UU        0.       0.       0.       1.
          7131.    1367.    30.00    1674.
             0.0      75.0
    97.5     256.       0.0       0.0      0.0      25.0
    2.30                          0.0     11.0      1.0    1.5317320320890
108 63AA0        0.       1.       0.       1.
          12356.   1120.    29.00    792.
             0.0      75.0
    97.0     150.       0.0       0.0      0.0      25.0
    1.20                          7.0     11.0      1.0    1.4370309398155
109 65AA0        0.       0.       0.       1.
          14271.   377.     14.00    982.
             0.0      75.0
    100.0    64.        0.0       0.0      0.0      25.0
    2.30                        131.0     21.0      1.0    1.7472191926755
110 63BAA        1.       0.       0.       1.
          3346.    1580.    49.00    1153.
             0.0      75.0
    23.0     500.       0.0       0.0      0.0      25.0
    1.30                          0.0      7.0      1.0    3.4157797048162
111 63CAA        1.       0.       0.       1.
          3346.    1580.    49.00    1153.
             0.0      75.0
    23.0     500.       0.0       0.0      0.0      25.0
    1.30                          0.0      7.0      1.0    2.7562336074586
112 65BAA        0.       1.       0.       1.
          3314.    1844.    29.00    1235.
             0.0      74.0
    97.5     90.        5.0       1.0      8.0      15.0
    1.20                          0.0     11.0      1.0    .4352216562293
113 61BBA        1.       0.       0.       1.
          5354.    1359.    49.00    1378.
             0.0      17.00
    70.0     380.       0.0       0.0      0.0       0.0
    1.30                          0.0     13.0      1.0    .2502321842071
114 65BAA        0.       1.       0.       1.
          3314.    1644.    29.00    1236.
             0.0      74.0
    97.5     90.        5.0       1.0      8.0      15.0
    1.20                         77.0     11.0      1.0    .9249355128164
115 63AF0        0.       1.       0.       1.
          4033.    1635.    51.00    1186.
             0.0      75.0
    23.0     502.       0.0       0.0      0.0      25.0
    1.20                          0.0     13.0      2.0    1.6363742652537
116 63AA0        0.       1.       0.       1.
          10712.   1120.    41.00    790.
             0.0      75.0
    97.0     150.       0.0       0.0      0.0      25.0
    1.20                          7.0     12.0      2.0    1.3389304438366
```

```
 N  LRU-ID  BOMBER  CARGO  SENSORY  COMM
        UP   V   W   CC
        DIG   AN   EM   FS   XMTR
SS   PD   BITFIT   IC   SRU   QPA   LSC/OH
UF

117 63121       0.      1.      0.      1.
            3345.    1581.    49.00   1153.
                0.0     75.0        0.0       0.0      21.0
      23.0     500.      0.0       0.0       7.0       1.0    1.6029917749853
      1.20
118 63AAA       0.      1.      0.      1.
            5355.    242.     9.00   1615.
                44.0    37.0        0.0       0.0      19.0
     100.0      35.      0.0      61.0      17.0       1.0     .4023131750404
      1.20
132 72AC0       1.      0.      0.      1.
           12332.     68.     4.00     62.
                0.0     75.0        0.0       0.0      25.0
      95.2      10.      0.0       0.0       2.0       1.0     .5076100034842
      1.30
```

APPENDIX B: PART 2

LISTING OF PHASE II DATA

```
N  LRU-ID  BOMBER  CARGO  SENSORY  COMM
      UP   V   W   CC
      DIG   AN   EM   FC   XMTR
SS  PD  BITFIT  IC  SRU  QPA  LSC/OH
UF

11 71FB0       0.       0.       0.     0.
         11308.      427.     13.90       61.
              0.0       33.0       67.0       0.0        0.0
 33.0    175.      17.0        1.         1.        1.   2.8097530626538
 2.30
27 729PD       0.       1.       0.     0.
        253330.     2471.     75.00     4503.
             30.0       59.0        1.0       0.0        0.0
 93.6    707.      4.0       371.        24.        1.   26.5312228617591
 1.29
36 51EA0       0.       0.       0.     0.
         13348.      542.     16.30     1031.
             87.0       7.5        0.0       5.5        0.0
100.0    131.      29.0      411.        14.        1.    .6406343738969
 2.30
37 52AA0       0.       0.       0.     0.
         15539.      668.     11.80     1443.
             96.4       0.0        1.0       4.5        0.0
100.0    205.       0.0       65.         7.        1.    .933467093767 22
 2.30
38 52AB0       0.       0.       0.     0.
         33115.      508.     11.60     1102.
            100.0       0.0        0.0       0.0        0.0
100.0    131.      44.0      125.        12.        1.   1.0125826841586
 2.30
39 63BD0       0.       0.       0.     0.
          2314.       78.      2.00       73.
              0.0      100.0        0.0       0.0        0.0
100.0      7.       0.0        7.         3.        1.    .5026535231998
 2.30
40 71AE0       0.       0.       0.     0.
        192215.     1700.     40.00     3193.
             85.7       7.1        1.5       5.0        0.0
 98.5    249.      41.0      946.         4.        1.   21.7330232595010
 2.30
41 71AK0       0.       0.       0.     0.
         13553.      353.      8.00      453.
             97.4       0.0        2.6       0.         0.0
 97.4     55.      21.0      191.         9.        1.   1.0168931037308
 2.30
42 74JA0       0.       0.       0.     0.
         23570.     8033.     21.50      263.
              0.0       63.4        0.0      36.6        0.0
 98.7     75.      54.0       39.        16.        1.   3.3712737350643
 2.30
43 74JC0       0.       0.       0.     0.
         25330.     1300.     21.00     1824.
             88.5       6.8        0.0       4.7        0.0
100.0   1003.       0.0      465.        16.        1.    .4137375916296
 2.30
```

```
N  LRU-ID  BOMBER  CARGO  SENSORY  COMM
     UP   V   W   CC
     DIG   AN   EM   PS   XMTR
SS   PD   BITFIT   IC   SRU   QPA   LSC/OH
UF

44 52GA1       0.       1.       0.       0.
          9431.     235.      7.00      273.
            47.0      47.0        6.0       0.0       0.0
     100.0      25.       3.6      63.0       7.0       1.0      .6306326352318
      3.13
45 71JCE       0.       1.       0.       0.
          7125.      51.      2.10       25.
            0.0       50.0       50.0       0.0       0.0
     50.0       7.       0.0       0.0       1.0       2.0      .1489706368900
      1.20
46 72AEJ       0.       1.       0.       0.
          80245.    2267.     58.00     4275.
            100.0       0.0       0.0       0.0       0.0
     100.0     333.       0.0    1154.0      93.0       1.0      2.3409342645327
      1.20
47 72CC0       0.       1.       0.       0.
          27391.    1097.     20.00     1289.
            70.0       30.0       0.0       0.0       0.0
     100.0      85.       0.0     275.0      37.0       1.0      .9450392555831
      1.20
48 71ZA0       0.       0.       0.       0.
          8127.      748.     26.50     2063.
            75.0       0.0       0.0       0.0      25.0
     99.9     100.       0.0     354.0       1.0       1.0      .1389487303896
      2.30
49 71ZB0       0.       0.       0.       0.
          1538.      154.      5.00      669.
            50.0       50.0       0.0       0.0       0.0
     100.0      25.       0.0      85.0       1.0       1.0      .0373146302798
      2.30
50 71ZD0       0.       1.       0.       0.
          516.       94.      2.00       94.
            0.0      100.0       0.0       0.0       .0
     100.0      10.       0.0      17.0       1.0       1.0      .0107370435367
      1.20
51 717A0       0.       0.       0.       0.
          8127.      748.     20.50     2063.
            75.0       0.0       0.0       0.0      25.0
     99.9     100.      16.0     354.0       1.0       1.0      .3328323075923
      2.30
52 717B0       0.       0.       0.       0.
          1538.      154.      5.00      569.
            50.0       50.0       0.0       0.0       0.0
     100.0      25.       0.0      65.0       1.0       1.0      .0303167224080
      2.30
53 71ZC0       0.       0.       0.       0.
          515.       94.      2.00       94.
            0.0      100.0       0.0       0.0       0.0
     100.0      10.       0.0      17.0       1.0       1.0      .6596989306556
      2.30
```

```
N  LRU-ID  BOMBER  CARGO  SENSORY  COMM
       UP   V   W   CC
       DIG   AN   EM   PG   XHTF
SS   PO   BITFIT   IC   SRU   QPA   LSC/OH
UF

54 73EG0       1.       0.       0.       0.
        65320.    1573.    47.40    3322.
           93.0      7.0         0.0        0.0        0.0
    100.0     225.      1.0   1543.0     21.0        2.0    5.3535953177258
    1.30
55 73EP0       0.       0.       0.       0.
       215330.    1435.    40.00    5015.
           80.0      20.0        0.0        0.0        0.0
    100.0     460.      1.0   1920.0     69.0        1.0    5.3140458227425
    2.30
56 73HA0       0.       0.       0.       0.
       340533.    3176.    70.00    2054.
           25.1      59.8       14.1        0.0        0.0
     85.5    1560.      0.0    113.0     47.0        1.0   11.9537025418060
    2.30
57 73HC0       1.       0.       0.       0.
       121411.    1027.    26.00    3027.
           41.7      39.0        0.0       19.3        0.0
    100.0     120.      2.0    713.0     22.0        1.0    2.8338628762542
    1.30
58 73NA0       0.       0.       0.       0.
        82553.    1156.    46.00     433.
            0.5      20.0       80.0        0.0        0.0
     20.0     287.      1.0      0.0     15.0        1.0    2.9490302675585
    2.30
59 73NB0       0.       0.       0.       0.
        77230.     467.    14.00    1813.
           85.0      10.0        1.0        0.        .0
    100.0     135.      0.0    543.0     11.0        1.0     .2350568896321
    2.30
60 73QB0       1.       0.       0.       0.
       105551.     659.    20.00     662.
            0.0      31.0        0.0       20.0        0.0
    100.0     128.      1.0      0.0     11.0        1.0    4.4339464882943
    2.30
61 73SC0       0.       0.       0.       0.
        86134.    1348.    15.60     994.
           90.0      10.0        0.0        0.0        0.0
    100.0      60.      0.0    611.0     13.0        1.0    2.7362040133779
    2.30
62 73KB0       0.       0.       0.       0.
        12754.    2755.    27.90     845.
            0.0     100.0        0.0        0.0        0.0
     98.5     290.      0.0      0.0     12.0        2.0    2.7155518394649
    2.30
63 73KE0       0.       0.       0.       0.
        35530.     713.    13.50     643.
            0.0       0.0        6.2       31.0        0.0
     31.8       9.      0.0      0.0     11.0        2.0     .8377926421405
    2.30
```

N LRU-ID BOMBER CARGO SENSORY COMM
UP V W CC
DIG AN EM PS XMTR
SS PD BITFIT IC SRU OPA LSC/OH
UF

64 73KF0 0. 0. 0. 0.
5380. 826. 27.60 272.
0.0 97.2 2.8 0.0 0.0
95.8 920. 2.0 0.0 3.0 2.0 .6308381204013
2.30
65 73KK0 0. 0. 0. 0.
2597. 622. 16.40 3164.
0.0 100.0 0.0 0.0 0.0
100.0 100. 3.0 7.0 14.0 2.0 2.2393565551840
2.30
67 73EG0 1. 0. 0. 0.
87249. 1573. 17.40 3322.
93.0 7.0 0.0 0.0 0.0
100.0 225. 0.0 1543.0 21.0 2.0 8.1766395002129
1.30
68 73HC0 1. 0. 0. 0.
121411. 1027. 26.00 3027.
41.7 39.0 19.3 0.0 0.0
100.0 120. 2.0 713.0 17.0 1.0 14.3548519875340
1.30
69 73LA0 1. 0. 0. 0.
31554. 1272. 36.00 2982.
0.0 100.0 0.0 0.0 0.0
99.1 290. 0.0 468.0 21.0 1.0 13.8516371801320
1.30
81 77EE0 1. 0. 1. 0.
335. 222. 6.50 9.
0.0 0.0 100.0 0.0 0.0
0.0 100. 0.0 0.0 3.0 1.0 .0349339993166
1.30
86 65BH0 0. 0. 1. 0.
2076. 750. 16.00 1308.
95.9 0.0 1.3 2.8 0.0
98.7 122. 0.0 578.0 7.0 1.0 .4535356982371
2.30
87 74FC0 0. 0. 1. 0.
125493. 1173. 25.70 965.
0.0 100.0 0.0 0.0 0.0
100.0 300. 63.0 39.0 6.0 1.0 3.9433311998746
2.30
88 74FJ0 0. 0. 1. 0.
32377. 1723. 26.20 1073.
0.0 91.3 8.7 0.0 0.0
91.3 123. 75.0 3.0 3.0 1.0 3.4393592035185
2.30
89 74FK0 0. 0. 1. 0.
4727. 119. 3.30 24.
90.0 0.0 10.0 0.0 0.0
90.0 20. 0.0 8.0 1.0 1.0 .0556344170000
2.30

```
N  LRU-ID  BOMBER  CARGO  SENSORY  COMM
       UP   V   W   CC
       DIG   AN   EM   PS   XMTR
SS   PD  BIT=IT   IC   SFU   QPA   LSC/OH
UF

 90 74F00       0.      0.      1.      0.
        120395.    1747.   160.00    8293.
            36.9       3.1        0.0       0.0       0.0
     130.0     537.      56.0  2392.0     31.0       1.0   7.9795675759528
      2.30
 91 74KA0       0.      0.      1.      0.
         56337.    2525.    38.00     522.
             0.0      33.9       0.0       6.1       0.0
      99.7     177.      30.0   181.0     19.0       1.0   1.8190920013107
      2.30
 92 74KC0       0.      0.      1.      0.
         36399.     592.    16.00    1755.
            92.2      17.5       0.0        .3       0.0
     100.0     130.       0.0  1375.0     18.0       1.0   1.0307230884290
      2.30
 93 75AE0       0.      0.      1.      0.
         34313.    1325.    37.00    4745.
            90.5       6.5       3.0       0.0       0.0
      97.0      58.       0.0   334.0     25.0       1.0    .0886597513855
      2.30
 94 74CA0       0.      0.      1.      0.
         19382.    1458.    38.00    2242.
            35.6      13.2       1.2       0.0       0.0
      98.8     517.       0.0    29.0     29.0       1.0    .1379087423807
      2.30
 95 74CB0       0.      0.      1.      0.
         22770.    1475.    44.00     819.
            98.9       0.0       1.1       0.0       0.0
      98.9     181.       0.0    38.0     19.0       1.0    .2022715619700
      2.30
 96 74CC0       0.      0.      1.      0.
         17325.     477.     5.00     917.
            99.1        .3        .6       0.0       0.0
      99.4     181.       1.0    41.0      1.0       1.0    .1566510027325
      2.30
 97 74FA1       0.      0.      1.      0.
         22395.    2701.    54.00    4193.
            99.8       0.0        .2       0.0        .0
      99.8     124.      20.0   869.0    101.0       1.0   1.9240397693109
      3.10
 98 74E90       0.      0.      1.      0.
         37137.     951.    19.00    1779.
            81.5      11.8        .7       6.0       0.0
      99.3     135.       0.0   345.0     17.0       1.0    .9069222908535
      2.30
 99 76AEA       1.      0.      1.      0.
         42930.    2666.    92.00    2375.
             0.0       0.0       0.0       0.0     100.0
      99.9    1300.      11.7   105.0     13.0       1.0  13.758953676559
       .30
```

```
N  LRU-ID  BOMBER  CARGO  SENSORY  COMM
        UP   V   W   CC
        DIG   AN   EM   PS   XMTR
SS   PD   BIT-IT   1C   SRU   QPA   LSC/OH
UF

100 73KA0       1.      0.      1.      0.
          10231.      622.     16.00     3241.
               0.0       90.8        1.2       0.0        0.0
      98.8      200.      10.0       0.0      13.0       2.0    4.5583732724246
      1.30
101 73PH0       0.      0.      1.      0.
          31318.      952.     32.00     1061.
               0.0       0.0        0.0      100.0        0.0
     100.0      465.       0.0       4.0      12.0       2.0    1.4339743589744
      1.20
102 73PB0       0.      0.      1.      0.
         555300.     1836.     51.00     3246.
              30.0      10.0        0.0       0.0        0.0
     100.0      500.       0.0     370.0      27.0       1.0    6.2395132307592
      1.20
103 73PD0       0.      0.      1.      0.
         139334.     4072.    145.00      725.
               0.0       0.0        0.0       0.0      100.0
     100.0     3000.       0.0       2.0       7.0       1.0   10.5358587521370
      1.20
104 73PF0       0.      0.      1.      0.
         234550.     2352.     51.00     2452.
              88.2      11.8        0.0       0.0        0.0
     100.0      275.       0.0    1095.0      34.0       1.0    1.1340384102564
      1.20
105 73PM0       0.      0.      1.      0.
          49533.     1445.     33.00      707.
              20.0      66.5       13.4       0.0        0.0
      85.1      145.       0.0       4.0       8.0       1.0    2.6533312820513
      1.20
119 55AL0       0.      1.      0.      1.
          38348.     1582.     40.00      361.
              58.7      11.9        0.0      19.0        0.0
     100.0      290.      44.0     290.0      23.0       1.0    1.7527493316663
      1.20
120 55AV0       0.      1.      0.      1.
         103000.     1549.     36.00      725.
              75.4       0.0        0.0      24.0        0.0
     100.0      300.      15.0     192.0      64.0       1.0    1.0590352431969
      1.20
121 61AA0       1.      0.      0.      1.
          24205.      363.     13.20      195.
               0.0      75.0        0.0       0.0       25.0
     100.0      150.        .5       7.0      17.0       2.0    1.5144458153403
      1.30
122 61AC0       1.      0.      0.      1.
          53550.     3370.     73.00      361.
               0.0      70.7       16.4      12.9        0.0
      63.6      150.        .5      69.0      14.0       2.0    1.6227537260350
      1.30
```

```
 N  LRU-ID  BOMBER  CARGO  SENSORY  COMM
        UP   V   W   CC
        DIG   AN   EM   PS   XMTR
SS  PD  BITFIT  IC  SRU  QPA  LSC/OH
UF

123 61AE0       0.      1.      0.      1.
          6958.     135.     4.30     471.
             0.0     100.0      0.0      0.0      0.0
     100.0     70.      0.0     33.0      7.0      2.0     .9344129138392
     1.20
124 62AA0       0.      1.      0.      1.
          3175.     364.    15.96    1116.
             0.0     75.0      0.0      0.0     25.0
     100.0    263.      0.0      0.0     15.0      2.0     .5327344115429
     1.20
125 63AG0       0.      0.      0.      1.
          7579.     428.    16.30     505.
            59.1     38.2      2.7      0.0      0.0
      97.3     32.      0.0     31.0      7.0      1.0     .4726744365490
     2.30
126 63BC0       0.      0.      0.      1.
          1335.     267.    12.00     742.
            53.1      7.6      0.0     39.3      0.0
     100.0     34.      0.0     73.0     12.0      1.0     .9372355726628
     2.30
127 53BF0       0.      0.      0.      1.
          2357.      78.     2.00     159.
            88.7      0.0      0.0     11.3      0.0
     100.0      3.      0.0     14.0      4.0      1.0     .0419141630584
     2.30
126 61AA0       1.      0.      0.      1.
         30391.     378.    13.12     195.
             0.0     75.0      0.0      0.0     25.0
     100.0    150.      0.0      7.0     13.0      1.0    2.9196324157021
     1.30
129 61AB0       1.      0.      0.      1.
         14526.     598.    23.13     432.
             0.0     43.8     32.0     24.2      0.0
     100.0    150.      0.0      3.0     13.0      1.0    2.7404372025860
     1.30
130 61AC0       1.      0.      0.      1.
         12353.     135.     4.17     471.
             0.0     100.0      0.0      0.0      .0
     100.0     70.     11.0     33.0      5.0      1.0    1.3702344055592
     1.30
131 72AA0       1.      0.      0.      1.
           556.      36.     1.00      33.
             0.0     100.0      0.0      0.0      0.0
     100.0      9.      0.0      0.0      1.0      1.0     .0442731928596
     1.30
133 64211       0.      1.      0.      1.
           333.     146.     4.00      37.
             0.0     100.0      0.0      0.0      0.0
     100.0      7.      0.0      9.0      1.0      1.0     .0555783291520
     1.20
```

```
 N  LRU-ID  BOMBER  CARGO  SENSORY  COMM
        UP   V   W   CC
        DIG   AN   EM   PS   XMTF
SS  PD  BITFIT  IC  SRU  QPA  LSC/OH
UF

134 64212       0.      1.      0.      1.
            575.      81.     2.40     133.
              0.0     91.9       8.1      0.0       0.0
     91.9        5.       0.0       0.0       1.0       1.0      .3144376476378
     1.20
```

APPENDIX C

ITEMIZED INPUT FOR AID *

* Extracted from McNichols [25]

APPENDIX C

Itemized Input for AID

1. <u>Title Card</u>

Card Column(s)	Use	Description
1	Card Type	Must contain the numeric value "1".
2-49	Job Title	Up to 48 alphabetic and/or numeric characters used to label the run.
50	IRUN	Numeric "0" for normal AID operation
51-56	NCPERM	Number of cases in the data file. May be omitted when data is from a disk or tape file.
79-80	IFMT	The number of cards used for the FORTRAN format statement (the next card or set of cards in the control card deck). Up to 4 cards may be used.

2. FORTRAN Foramt Card(s)

Card Column(s)	Use	Description
1-78	Data Format	FORTRAN format statement beginning with a left parenthesis and ending with a right parenthesis. Only integer fields of the form: Iw, where w is the number of characters used to describe a variable, can be specified. The characters: X can be used to skip columns, T to tab to a desired character position, and / to indicate the beginning of a new record for multiple record cases. <u>Warning:</u> be careful not to extend the format statement beyond column 78 as these characters are not processed by AID. If more than 78 characters are needed for the format statement, use another format card and change the count in column 80 of the title card.

APPENDIX C

Itemized Input for AID (cont'd)

3. Description Card

Card Column(s)	Use	Description
1	Card Type	Must contain the numeric value "3"
2-6	Stopping Rule:P1	Minimum value of TSS_i/TSS_T to consider group i for splitting. (Section 8.2.2, paragraph 2). A decimal point is implied to the left of col. 2.
7-11	Stopping Rule:P2	Minimum value of BSS_i/TSS_T to permit group i to split (Section 8.2.2, paragraph 3). A decimal point is implied to the left of column 7.
12-16	Stopping Rule MAXGP	Maximum number of subgroups into which the set of data will be split.
17-21	Stopping Rule: NMIN	Minimum number of observations which must be in a group after it is split. Value must be at least 2.
22-26	Iteration Print: KSTOP	Number of AID iterations for which detailed information will be printed. Only summary results for iterations will be output after this point.
27-29	No. of Variables NV	Specifies number of variables to be read from each case. This will be the total number of variables described by the format statement.
33	Rewind: KRW	Should be the numeric value "1" if input data is on a disk or tape file, left blank otherwise.
34	Missing Values: IOPT	Set to "1" if a case with any out-of-range predictor values is to be rejected, blank or zero otherwise. The "1" value is analogous to listwise deletion in SPSS, as far as the predictor variables are concerned. There is no capability in AID which corresponds directly to a pairwise deletion option. The IOPT setting must be considered when predictor cards (type 4) are coded.

APPENDIX C

Itemized Input for AID (cont'd)

Card Column(s)	Use	Description
37	Input Medium: ICARD	If zero or blank, the data file is assumed to be a disk or tape file with the local file name "TAPE25". If set to "1", data is assumed to be on punched cards which follow the AID control cards.
38	Tree Control: ITREE	This parameter controls the output of computer printed tree diagrams summarizing the splits. If set to zero or blank, no diagrams are generated. If set to "1", only a detailed tree is generated. If set to "2", both a detailed and a skeleton tree will be produced.

4. Predictor Card(s)

Card Column(s)	Use	Description
1	Card Type	Must contain the numeric value "4" There will be one predictor card for each predictor variable to be used in the AID run. However, all predictors described by the format statement do not need to be used in the AID run. The NV parameter (card 3) has a value associated with the number of variables described by the format statement, not the number of predictor cards used in the run.
2-19	Predictor Name	Up to 18 alphabetic or numeric characters used to label the predictors in the AID output.
20-22	Field Number	A variable number which must correspond to the variable sequence provided by the format statement. This is, the third variable described by the format statement represents field number 3 for predictor variable numbering purposes
23	Predictor Type: KBL1	Zero or blank for predictors to be treated as nominally scaled, "1" for variables to be treated as ordinally scaled. The example in section 8.1 illustrates the nature of the treatment of nominal and ordinal variables in AID.

AD-A080 407 AIR FORCE INST OF TECH WRIGHT-PATTERSON AFB OH SCHOO--ETC F/G 12/1
CROSS VALIDATION OF SELECTION OF VARIABLES IN MULTIPLE REGRESSI--ETC(U)
DEC 79 J R CAFARELLA
UNCLASSIFIED AFIT/GOR/MA/79D-2 NL

2 OF 2
AD
A080407

END
DATE
FILMED
3-80
DDC

APPENDIX C

Itemized Input for AID (cont'd)

Card Column(s)	Use	Description
24	Predictor Definition: KBL2	This parameter, used in conjunction with the IOPT value on card 3, tells AID how to interpret the values on the remainder of the predictor card. A zero value indicates that the range of possible values for this predictor variable will be divided into intervals of fixed length. A value of "1" means that the range of values for this predictor will be divided into intervals of varying length. When KBL2 is set to zero, minimum and maximum values and an interval length will be provided. When KBL2 is set to "1", boundaries for the intervals into which the range of predictor values will be divided will be specified. Figure 8.7 summarizes the interpretation of IOPT/KBL2 value combinations and should be referenced in choosing the desired values and predictor card format.

A. IOPT Equal Zero and KBL2 Equal Zero:

Card Column(s)	Use	Description
25-30	Minimum Predictor Value:MIN	Predictor variable values less than or equal to this value will be recoded to an internal value (recode category) of 00.
31-36	Maximum Predictor Value:MAX	Predictor variable values greater than or equal to this value will be recoded to the highest recode category value used for this predictor.

APPENDIX C

Itemized Input for AID (cont'd)

Card Column(s)	Use	Description
37-42	Interval Length: INT	The length of the range of values for this predictor to be recoded into a single recode category. The recode

- **IOPT**
  - Blank/Zero: Retain Cases With Out of Range Predictor Values
    - **KBL2**
      - Blank/Zero: Recode Equal Intervals
        - Predictor Card Provides:
          - a) Minimum value of predictor--MIN
          - b) Maximum value of predictor--MAX
          - c) Interval length--INT
          - Note: Number of intervals: (MAX-MIN)/INT + 1
          - Values $\leq$ MIN recode to 00
          - Values $\geq$ MAX recode to highest recode category
      - One: Specified Recode Categories
        - Predictor Card Provides:
          - a) Upper boundary for each predictor value range
          - b) Recode values (00 to 39) for each predictor value range
  - One: Reject Cases With Out of Range Predictor Values
    - **KBL2**
      - Blank/Zero: Recode Equal Intervals
        - Predictor Card Provides:
          - a) Minimum value of predictor--MIN
          - b) Maximum value of predictor--MAX
          - c) Interval length--INT
          - Note: Number of intervals: (MAX-MIN)/INT
          - Values <MIN or >MAX lead to rejection of case
      - One: Specified Recode Categories
        - Predictor Card Provides:
          - a) Minimum value of predictor variable
          - b) Upper boundary for each range of predictor values
          - c) Recode values (00 to 39) for each range of predictor values

Predictor Card Coding: Interpretation of IOPT/KBL2 Values

Card Column(s)	Use	Description
		category assigned to a specific predictor value between the MIN and MAX value will be:

$$\text{Recode Category} = \frac{\text{Predictor Value-MIN}}{\text{INT}}$$

The number of recode categories will be:

$$\text{NCAT} = \frac{\text{MAX-MIN}}{\text{INT}} + 1$$

The highest numbered recode category will be NCAT-1, and values greater than or equal to MAX will be assigned this value.

Card Column(s)	Use	Description
44-45 53-54 62-63		In a basic application of AID, each of these pairs of columns should contain the value "-1". These columns can be used in conjunction with other predictor card parameters to alter the recoding process by assigning specific recode categories to specific numeric values of the predictor variable. Since this is a less often used capability, it will not be discussed in detail here.

B. IOPT Equal Zero and KBL2 Equal One:

Card Column(s)	Use	Description
25-27	Lowest Recode Category	A value between 00 and 39 which is the numeric value to be used internally by AID to represent predictor variable values less-than-or-equal-to the first specified input value.
28-33	First Specified Input Value	A value of the predictor variable--used with lowest recode category.
34-36	Second Recode Category	A value between 00 and 39 which is the value to be used internally by AID to represent predictor variable values strictly greater than the first specified input value, and less-than-or-equal-to the second specified input value.

APPENDIX C

Itemized Input for AID (cont'd)

Card Column(s)	Use	Description
37-42	Second Specified Input Value	A value of the predictor variable-- used in conjunction with the seond recode category as the boundary of the predictor variable values to be recoded to the value specified by the second recode category.
43-45	Third Recode Category	A value between 00 and 39 which is the value to be used internally by AID to represent predictor variable values strictly greater than the second specified input value and less-than-or-equal-to the third specified input value.
46-51	Third Specified Input Value	A value of the predictor variable--used in conjunction with the third recode category.
52-54 61-63	Fourth & Fifth Recode Categories	The descriptions of these field are comparable to those given for the first, second and third recode categories and specified input values.

C. IOPT Equal one and KBL2 Equal Zero:

25-30	Minimum Predictor Value:MIN	Predictor variable values strictly less than this value will cause the case to be rejected.
31-36	Maximum Predictor Value:MAX	Predictor variable values greater than or equal to this value will cause the case to be rejected.
37-42	Interval Length: INT	The length of the range of values for this variable to be recoded into a single recode category. The recode category assigned to a specific predictor variable between MIN and MAX will be: $\text{Recode Category} = \dfrac{\text{Predictor Value-MIN}}{\text{INT}}$

APPENDIX C

Itemized Input for AID (cont'd)

Card Column(s)	Use	Description
		The number of recode categories will be: $$NCAT = \frac{MAX-MIN}{INT}$$
44-45 53-54 62-63		In a basic application of AID, each of these pairs of columns should contain the value"-1". These columns can be used in conjunction with other predictor card parameters to alter the recoding process by assigning specific recode categories to specific numeric values of the predictor variable. Since this is a less often used capability, it will not be discussed in detail here.

D. IOPT Equal One and KBL2 Equal One:

Card Column(s)	Use	Description
25-27	Recode Category	Used only when more than one predictor card is required to describe the predictor variable. On the first predictor card for a variable this field should be blank.
28-33	First Specified Input Value	Predictor variable values less-than-or-equal-to this value will cause the case to be rejected.
34-36	Second Recode Category	A value between 00 and 39 which is the value to be used internally by AID to represent predictor variable values strictly greater than the first specified input value, and less-than-or-equal-to the second specified input value.
37-42	Second Specified Input Value	A value of the predictor variable associated with the second recode category.
43-45	Third Recode Category	A value between 00 and 39 which is the value to be used internally by AID to represent predictor variable values strictly greater than the second specified input value and less-than-or-equal-to the third specified input value.

APPENDIX C

Itemized Input for AID (cont'd)

Card Column(s)	Use	Description
46-51	Third Specified Input Value	A value of the predictor variable associated with the third recode category.
52-54 61-63	Fourth & Fifth Recode Categories	The descriptions of these fields are comparable to those given for the first, second, and third recode categories and specified input values.

5. Criterion Card

Card Column(s)	Use	Description
1	Card Type	Must contain the numeric value "5"
2-19	Criterion Name	Up to 18 Alphabetic or numeric characters used to label the criterion variable in the AID output.
20-22	Field Number	A variable number which must correspond to the variable sequence provided by the format statement. That is, the third variable described by the format statement represents field number 3 for criterion variable identification purposes. The criterion variable does not have to be the field which is physically last in each case as long as the proper field numbers are used to identify predictors and the criterion.
23-24	Weight Field	A variable number representing a weight field in each case, used to weight the values in AID computations. This field can be left blank, causing all cases to be equally weighted, and this is the normal mode of operation.
25-30	Maximum Criterion Value: YMAX	If the criterion variable value is strictly greater than YMAX in a case, the case is rejected. Values up to "999999" can be specified for YMAX.

Itemized Input for AID (cont'd)

Card Column(s)	Use	Description
31-36	Minimum Criterion Value: YMIN	If the criterion variable value is strictly less than YMIN in a case, the case is rejected.
37-42 43-48	Deletion Values: MD1,MD2	If the criterion variable value is equal to either of these values, the case is rejected. If the use of deletion values is not desired, or only one deletion value is desired, setting MD1 and/or MD2 to values outside the range of YMIN to YMAX deactivates their use.

6. AID End-Of-Job Card

1	Card Type	Must contain the numeric value "9" Indicates the end of all of the AID control cards.

APPENDIX D

SELECTED AID OUTPUT

```
UNIVERSITY OF TEXAS VERSION OF AID4 AS ADOPTED FROM AFHRL,LACKLAND AFB
MODIFIED FOR SDCHSCRG BY CAPT LL GOOCH, OCT 72

ONTROL CARD PARAMETERS

TITLE CARD

CARD
TYPE                 TITLE                             IRUN  NOPERM  NREELS  NSAMA  NSAMB  NUM  IFMT
 1   AID RUN ON TEST DATA/// Y=LSC/OH                   1      63      0       0      0     0    2

FORMAT CARD OR CARDS

(2X,3X,1X,5X,7(?X,I1)/11X,I?,2X,I?,4X,I?,2X,I6/13X,2(I?,3X),1X,2(I5,3X),I5,2X/
12X,I5,2X,I5,3X,I6,2X,3(I?,2X)/5X,I5)

DESCRIPTION CARD

CARD
TYPE   P1      P2     MAXGP  NMIN  KSTOP  NV  NR  KRW  IOPT  KREJ  IRESID  ICARD  ITREE  NOGO  IESS  IRSQ1  IRSQ2
 3   .01500  .05000    30     3     1    23   0   1    0     0      0       0      2     0     0      0      0
```

VARIABLE		FIELD NO.	RECODE	CORRESPONDS TO	TYPE
1	FGTNAV	1	0	LESS THAN 1	FREE-FLOATING
			1	1 OR OVER	
2	BOMNAV	2	0	LESS THAN 1	FREE-FLOATING
			1	1 OR OVER	
3	CARNAV	3	0	LESS THAN 1	FREE-FLOATING
			1	1 OR OVER	
4	FGTSEN	4	0	LESS THAN 1	FREE-FLOATING
			1	1 OR OVER	
5	BOMSEN	5	0	LESS THAN 1	FREE-FLOATING
			1	1 OR OVER	
6	FGTCOM	6	0	LESS THAN 1	FREE-FLOATING
			1	1 OR OVER	
7	BOMCOM	7	0	LESS THAN 1	FREE-FLOATING
			1	1 OR OVER	
8	UNITPRICE	8	0	LT. OR EQ. TO 2241	FREE-FLOATING
			1	2242 TO 3915	
			2	3916 TO 6410	
			3	6411 TO 19274	
			4	19275 OR OVER	
9	VOLUME	9	0	LT. OR EQ. TO 275	FREE-FLOATING
			1	276 TO 560	
			2	561 TO 1377	
			3	1378 TO 1734	
			4	1735 OR OVER	
10	WEIGHT	10	0	LT. OR EQ. TO 850	FREE-FLOATING
			1	851 TO 1500	
			2	1501 TO 3500	
			3	3501 TO 4900	
			4	4901 OR OVER	
11	COMPONENTCOUNT	11	0	LT. OR EQ. TO 82	FREE-FLOATING
			1	83 TO 399	
			2	400 TO 911	
			3	912 TO 1186	
			4	1187 OR OVER	

TRY ON PREDICTOR 19 BITFIT

CODE	N	TOTAL WEIGHT	SUM OF Y	SUM Y-SQUARE	MEAN	STD. DEV.	B S S	W S S
1	7	7.000000	26911.000	.17682761E+09	3844.4286	3281.3366		
							.76304044E+09	.15107704E+11
2	7	7.000000	63353.000	.16769368E+10	9050.0000	12563.190		
							.94382711E+09	.14926917E+11
0	43	43.000000	669728.00	.23269863E+11	15575.070	17272.645		
							75378045.	.15795366E+11
3	6	6.000000	102330.00	.25567474E+10	17055.333	11623.412		

MAX. BSS= .94382711E+09 BSS/TSS = .05947 BETWEEN CODES 1 2
AND CODES 0 3

TRY ON PREDICTOR 20 IC

CODE	N	TOTAL WEIGHT	SUM OF Y	SUM Y-SQUARE	MEAN	STD. DEV.	B S S	W S S
2	7	7.000000	74682.000	.69413264E+09	10697.429	3006.6057		
							70425714.	.15600319E+11
1	5	5.000000	67305.000	.15509333E+10	13461.200	11353.705		
							50126813.	.15620618E+11
0	44	44.000000	609951.00	.22657108E+11	13630.932	17985.564		
							39967096.	.15630757E+11
3	7	7.000000	111399.00	.25792113E+10	159 1.286	10692.716		

MAX. BSS= 70425714. BSS/TSS = .00444 BETWEEN CODES 2
AND CODES 1 0 3

TRY ON PREDICTOR 21 SRU

CODE	N	TOTAL WEIGHT	SUM OF Y	SUM Y-SQUARE	MEAN	STD. DEV.	B S S	W S S
0	13	13.000000	38018.000	.25213437E+09	2923.6923	3293.4745		
							.18979223E+10	.13972622E+11
2	11	11.000000	165735.00	.35520373E+10	15066.816	9793.1234		
							.10477609E+10	.14822963E+11
1	12	12.000000	184719.00	.63416567E+10	15393.250	17073.964		
							.70512208E+09	.15165622E+11
3	14	14.000000	228417.00	.61572216E+10	16315.500	12902.060		
							.44181826E+09	.15428926E+11
4	13	13.000000	219159.00	.11471203E+11	16851.462	22932.406		

MAX. BSS= .18979223E+10 BSS/TSS = .11959 BETWEEN CODES 0
AND CODES 2 1 3 4

TRY ON PREDICTOR 22 OPA

CODE	N	TOTAL WEIGHT	SUM OF Y	SUM Y-SQUARE	MEAN	STD. DEV.	B S S	W S S
0	58	58.000000	752679.00	.24619599E+11	12980.672	15993.366		
							.36552979E+09	.15505215E+11
1	5	5.000000	109459.00	.30566161E+10	21891.800	11475.117		

MAX. BSS= .36552979E+09 BSS/TSS = .02303 BETWEEN CODES 0
AND CODES 1

SPLIT GROUP 18 ON PREDICTOR 10 WEIGHT INTO GROUP 22 WITH CODES 0
AND GROUP 23 WITH CODES 1 3

BSS = 9361.100. BSS/TSS = .43197 T-VALUE 3.93

CURRENT SUMMARY

NCF	TOTAL TSS	TOTAL BSS	TOTAL WSS	R-SQUARED	R	F-RSQ	DF1	DF2	F-ANOVA	DF1	DF2
12	.15679744E+11	.11022106E+11	.46485981E+10	.69449712	.83336	.9846	1	51	10.5398	11	51

GROUP 22 CANNOT BE SPLIT FOR FAILURE TO CONTAIN P1= .005010
AS SPECIFIED BY THE USER ON CARD 3, COL. 2-6.

PROPORTION OF THE TOTAL SUM OF SQUARES
TSS IN THIS GROUP= .54002651E+07
TSS= .15679744E+11
P1*TSS = .78353722E+08

GROUP	N	TOTAL WEIGHT	SUM OF Y	SUM Y-SQUARE	T S S	MEAN	STD. DEV.
22	12	12.000000	11322.00	16162532.	5400265.6	943.50000	670.63610

CANDIDATE GROUPS ARE AS FOLLOWS.

GROUP	N	TOTAL WEIGHT	SUM OF Y	SUM Y-SQUARE	T S S	MEAN	STD. DEV.
23	13	13.000000	62617.000	.43947[illegible]E+09	.13385923E+09	4816.6923	3208.9805

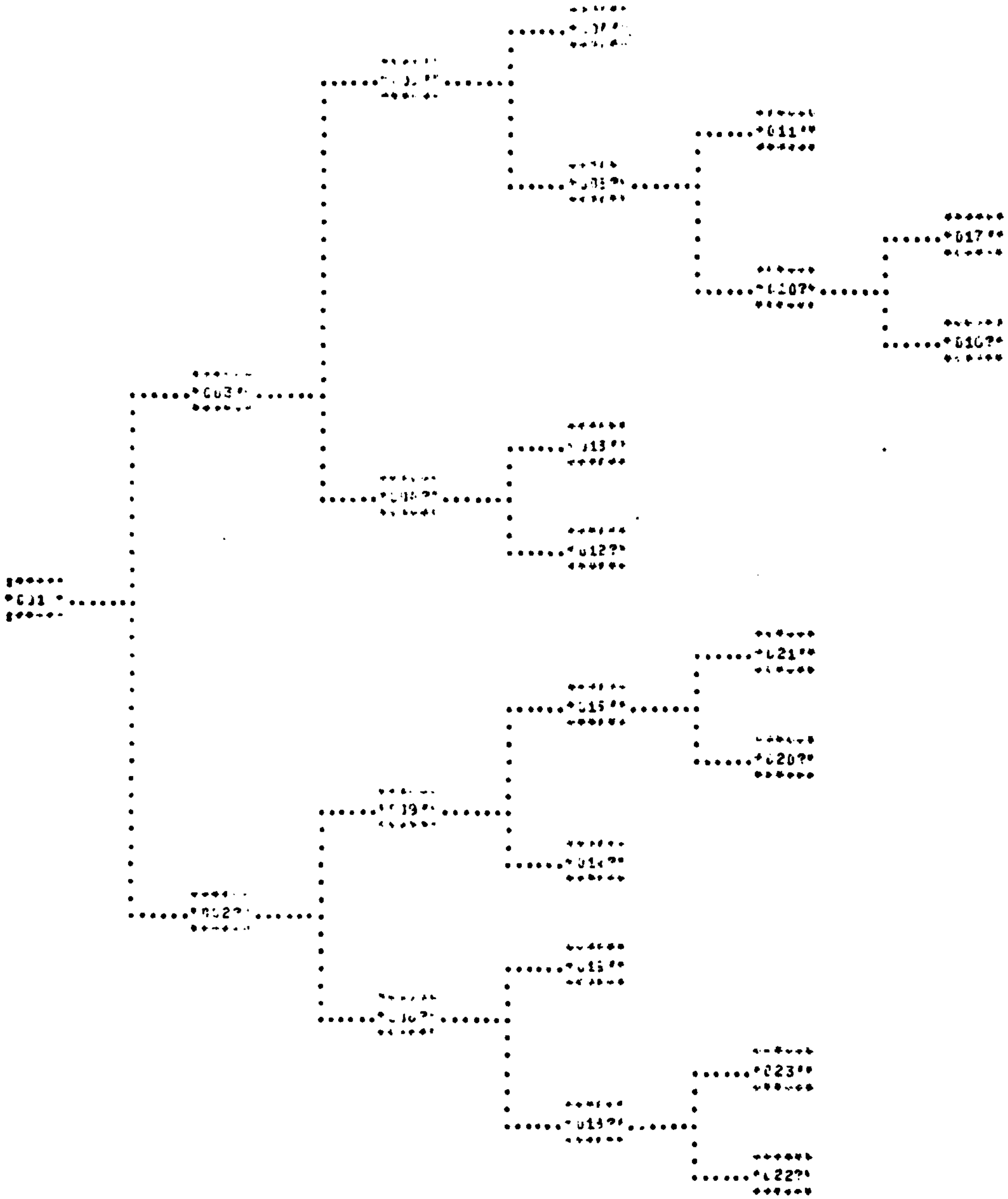
SUMMARY TABLE 22 VARIABLES, NO OF SUBGROUPS IS 12 GRAND MEAN = .263840 ROUND CRIT. = .010TSST =********

	SUB(VARIANT) SUMSSS/TSST	RECONSTRUCTED SUMBSS/TSST	S D	COEFF VAR	VARIANCE PCT	GRD MEAN	N GROUPS	RANK
1	.11965	.02359	.00493	2.04971	.00002	11.57082	10.	20.0
2	.10515	.13193	.02713	2.40576	.00074	64.25786	10.	13.0
3	.04297	.04393	.00675	1.76364	.00005	22.51727	11.	19.0
4	.07522	.07322	.00972	1.49179	.00009	38.37130	12.	15.0
5	.01170	.02106	.00409	2.44945	.00002	9.83934	7.	22.0
6	.04316	.10359	.01335	1.54622	.00018	50.81842	5.	14.0
7	.01513	.02270	.00483	2.57470	.00002	11.13634	8.	21.0
8	.33432	.33432	.04295	1.45506	.00185	173.82367	12.	5.0
9	.37079	.37679	.07953	2.53264	.00632	164.84655	12.	4.0
10	.47772	.57325	.07534	1.06079	.00029	261.23033	10.	1.0
11	.43605	.45586	.04275	1.24838	.00179	199.59828	12.	3.0
12	.05037	.05337	.01193	2.26613	.00012	28.67942	12.	17.0
13	.45099	.45099	.06652	1.77011	.00443	221.24623	12.	2.0
14	.13045	.19054	.01895	1.25935	.00036	69.56823	10.	10.0
15	.06134	.07361	.01535	2.00176	.00024	36.11201	10.	16.0
16	.28700	.25700	.04453	1.86852	.00200	140.79610	12.	9.0
17	.30329	.31329	.03332	1.31819	.00111	143.76599	12.	7.0
18	.31477	.31477	.03793	1.49781	.00144	154.41938	12.	6.0
19	.14601	.14601	.01725	1.41744	.00030	71.62794	12.	12.0
20	.17080	.17980	.01734	1.16348	.00030	67.71805	12.	11.0
21	.30181	.30181	.03571	1.41999	.00128	149.06215	12.	8.0
22	.04516	.05254	.00933	2.12377	.00009	25.77232	11.	18.0

ANALYSIS WITH   ?SS/TSS(I)

TRIAL/GRP		19	20	21	22
1	1	.059470	[illegible]	.119586	.123032
2	3	.020314	.072371	.155911	0.000000
3	5	.081547	.155541	.067459	0.000000
4	2	.080336	.224324	.297557	.194102
5	6	.343290	.343290	.147103	0.000000
6	4	0.000000	0.000000	.209705	0.000000
7	9	0.000000	.252560	.163351	0.000000
8	10	.253718	.253718	.062249	0.000000
9	8	.032436	.237549	.242166	0.000000
10	15	0.000000	0.000000	0.000000	0.000000
11	13	.315454	0.000000	.237350	0.000000
12	23	.039271	0.000000	.337336	0.000000
SUBSUM		.92566	1.52761	1.93713	.21713
TRIALS		12.	12.	12.	11.
MEANBS		.07714	.12730	.16143	.01974
S D		.10434	.12347	.09814	.05553
CF VAR		1.35262	.96934	.61229	2.81317
VAR		.01089	.01525	.00977	.00308
RECSUM		.92566	1.52761	1.93713	.23667

TREE STRUCTURE OF FIRST EIGHT LEVELS, GROUP NUMBER IS GIVEN IN CENTER OF THE BOX, THE PAIR MEMBER WITH THE HIGHER MEAN IS ON TOP



```
GROUP   1          MEAN= 136?7.50
 N=    53          S.D.= 15?71.??
        LEVEL 1

GROUP   3          MEAN= ?????.?? RSQ =  .294
 N=    27          S.D.= 19133.13 PROB= 0.00?
 PREDICTOR  9             VOLUME
 CODES  4  3

GROUP   2          MEAN= ?????.97 RSQ =  .294
 N=    36          S.D.= ????.?? PROB= 0.00?
 PREDICTOR  9             VOLUME
 CODES  4  1  2

GROUP   5          MEAN= 31399.?4 RSQ =  .435
 N=    18          S.D.= 19640.58 PROB=  .000
 PREDICTOR 13         PERCENTANALOG
 CODES  5  2

GROUP   4          MEAN= 1166?.33 RSQ =  .435
 N=     9          S.D.=  9052.37 PROB=  .000
 PREDICTOR 13         PERCENTANALOG
 CODES  3  1  4

GROUP   9          MEAN= 15300.?7 RSQ =  .595
 N=     9          S.D.=  646?.32 PROB=  .0?7
 PREDICTOR 13               WEIGHT
 CODES  2

GROUP   6          MEAN=  404?.72 RSQ =  .595
 N=    29          S.D.=  43??.?2 PROB=  .0?7
 PREDICTOR 13               WEIGHT
 CODES  0  1  4  3

GROUP   7*FINAL MEAN= 49079.53 RSQ =  .540
 N=     4       S.D.= 29?40.9? PROB=  .002
 PREDICTOR 11     COMPONENTCOUNT
 CODES  2  ?

GROUP   6       MEAN= 25506.06 RSQ =  .540
 N=    12       S.D.=  9193.76 PROB=  .001
 PREDICTOR 11     COMPONENTCOUNT
 CODES  4  3  1

GROUP  13*FINAL MEAN= 18340.75 RSQ =  .638
 N=     4       S.D.=  9629.56 PROB=  .083
 PREDICTOR 11     COMPONENTCOUNT
 CODES  1  3

GROUP  12*FINAL MEAN=  6328.60 RSQ =  .636
 N=     5       S.D.=  3030.33 PROB=  .083
 PREDICTOR 11     COMPONENTCOUNT
 CODES  4  2

GROUP  15       MEAN= 19863.17 RSQ =  .661
 N=     6       S.D.=  6763.99 PROB=  .059
 PREDICTOR 13      PERCENTANALOG
 CODES  2  5  1

GROUP  14*FINAL MEAN=  6313.67 RSQ =  .661
 N=     3       S.D.=  1365.32 PROB=  .059
 PREDICTOR 13      PERCENTANALOG
 CODES  4  3

GROUP  19*FINAL MEAN= 19616.50 RSQ =  .684
 N=     4       S.D.=  4620.07 PROB=  .140
 PREDICTOR 16        PERCENTXMTR
 CODES  3  2

GROUP  18       MEAN=  2957.56 RSQ =  .684
 N=    25       S.D.=  3052.07 PROB=  .140
 PREDICTOR 16        PERCENTXMTR
 CODES  0  1
```

APPENDIX E

SELECTED SPSS OUTPUT

```
                                                                12/03/79     15.30.20.     PAGE    1
VOGELBACK COMPUTING CENTER
NORTHWESTERN UNIVERSITY

S P S S - - STATISTICAL PACKAGE FOR THE SOCIAL SCIENCES

VERSION 7.0 -- JUNE 27 1977


          RUN NAME        REGRESSION  LOGNEWDATA
          PRINT BACK      CONTROL
          VARIABLE LIST   N,ID,X1 TO X20
          INPUT MEDIUM    DISK
          N OF CASES      UNKNOWN
          INPUT FORMAT    (2X,F3.0,1X,A6,12(1X,F2.0),1X,F4.0,F16.13/
                          3(4X,F16.13)/3(4X,F16.13))

THE INPUT FORMAT PROVIDES FOR  22 VARIABLES.   22 WILL BE READ
IT PROVIDES FOR  3 RECORDS (*CARDS*) PER CASE.  A MAXIMUM OF   69 *COLUMNS* ARE USED ON A RECORD.

WARNING - A NUMERIC VARIABLE HAS A WIDTH GREATER THAN 14.  SMALL ROUNDING/TRUNCATION ERRORS MAY OCCUR.
          COMPUTE         X21=X1*X15
          COMPUTE         X22=X1*X18
          COMPUTE         X23=X2*X15
          COMPUTE         X24=X2*X16
          COMPUTE         X25=X2*X17
          COMPUTE         X26=X8*X16
          COMPUTE         X27=X8*X17
          COMPUTE         X28=X9*X15
          COMPUTE         X29=X9*X17
          COMPUTE         X30=X10*X16
          COMPUTE         X31=X12*X18
          COMPUTE         X32=X12*X14
          COMPUTE         X33=X13*X17
          COMPUTE         X34=X13*X14
          REGRESSION      VARIABLES=X1 TO X34/
                          REGRESSION=X20 (23,1.0,0.0) WITH X2,X4,X5,X8,X10,X11,X14,
                          X15,X17,X21 TO X34 (2)
          STATISTICS      ALL
          READ INPUT DATA
          FINISH

   00053600 CM NEEDED FOR REGRESSION


   OPTION - 1
   IGNORE MISSING VALUE INDICATORS
```

FILE NONAME (CREATION DATE = 12/03/79)

* M U L T I P L E   R E G R E S S I O N *

VARIABLE	MEAN	STANDARD DEV	CASES
X1	.3521	.4810	71
X2	.0423	.2026	71
X3	.0986	.3002	71
X4	.2535	.4351	71
X5	.0423	.2026	71
X6	.0423	.2026	71
X7	.0563	.2322	71
X8	.5634	.4935	71
X9	.7606	.4298	71
X10	.4085	.4950	71
X11	.2676	.4459	71
X12	.0986	.3002	71
X13	.4085	.4950	71
X14	.0045	.0006	71
X15	9.8685	1.7144	71
X16	6.5459	1.1564	71
X17	2.8689	1.1108	71
X18	6.5604	1.5958	71
X19	4.7257	1.4746	71
X20	.0200	1.7993	71
X21	3.5376	4.9273	71
X22	2.3628	3.3086	71
X23	.4711	2.2618	71
X24	.3020	1.4486	71
X25	.1507	.7244	71
X26	3.8110	3.4463	71
X27	1.7435	1.6950	71
X28	7.5424	4.5430	71
X29	2.1387	1.5467	71
X30	2.6942	3.3264	71
X31	.6646	2.0485	71
X32	.0805	.0014	71
X33	1.3449	1.7064	71
X34	.0018	.0022	71

DEPENDENT VARIABLE.. X20

PARAMETERS.. MAXIMUM STEP 23.. F TO ENTER 1.000000.. TOLERANCE .000000.. F TO REMOVE .005000

MEAN RESPONSE .02000 STD. DEV. 1.79825

VARIABLE(S) ENTERED ON STEP NUMBER 1.. X2
X30
X34
X5
X11
X27
X28
X4
X32
X14
X22
X15
X17
X8
X33
X29
X21
X10
X31
X25
X26
X23

MULTIPLE R	.88320
R SQUARE	.78004
ADJUSTED R SQUARE	.67923
STD DEVIATION	1.01847

ANALYSIS OF VARIANCE	DF	SUM OF SQUARES	MEAN SQUARE	F	SIGNIFICANCE
REGRESSION	22.	176.57021	8.02592	7.73752	.000
RESIDUAL	48.	49.78910	1.03727		
COEFF OF VARIABILITY	5092.2 PCT				

------------------- VARIABLES IN THE EQUATION -------------------

VARIABLE	B	STD ERROR B	F SIGNIFICANCE	BETA ELASTICITY
X2	9.2856618	17.346404	.22377333 .638	.9244860 17.33545
X30	-.35382193	.25552978	1.9172041 .173	-.6545027 -47.66247
X34	59.533507	207.10304	.82632328E-01 .775	.0740860 5.46818
X5	.99081459	.83450842	1.4074136 .241	.1115394 2.09152
X11	.12680075	.31751393	.15948436 .691	.0314392 1.69658
X27	.49276704	.61187387	.64857393 .425	.4644831 42.95697

---------- VARIABLES NOT IN THE EQUATION ----------

VARIABLE	PARTIAL	TOLERANCE	F SIGNIFICANCE
X24	-1.00000	-.00008	0 1.000

REGRESSION LOGNEWDATA

X28	.64676379E-01	.13817898	.21988237 .642	.1633948 24.39024
X4	-.33287310	.39887747	.69643147 .488	-.0811087 -4.21941
X32	531.35247	667.96390	.63278973 .430	.4083768 12.05731
X14	1102.0708	280.38031	15.458658 .000	.3622141 246.66791
X22	-.13099197	.22804494	.32995024 .568	-.2410118 -15.47528
X15	.36000615	.17946696	4.0239337 .051	.3432206 177.63222
X17	.60315963	.43885436	1.8889640 .176	.3725769 86.51779
X8	-1.3736140	2.5219780	.29665167 .589	-.3815462 -38.69232
X33	.55957127E-01	.28894300	.37504765E-01 .847	.0531003 3.76268
X29	-.13646127	.43869986	.96757322E-01 .757	-.1173699 -14.59241
X21	.13008302	.15751183	.68204793 .413	.3564363 23.00848
X10	2.6000225	1.6961873	2.3496684 .132	.7157682 53.09764
X31	-.36370529	.46191312	.61998136 .435	-.4143276 -12.08513
X25	-.75005236	2.6271934	.81507780E-01 .776	-.3021493 -5.65178
X26	-.14280299	.60250900	.56175572E-01 .814	-.2735781 -27.21048
X23	-.34773058	1.1390687	.93193562E-01 .761	-.4373705 -8.13008
(CONSTANT)	-10.438349	1.4928598	48.890528 .000	

NOTE- 1 VARIABLES WERE NOT FORCED DUE TO INSUFFICIENT TOLERANCE.
INCLUSION LEVELS WERE SET TO ZERO.

ALL VARIABLES ARE IN THE EQUATION.

FILE NONAME (CREATION DATE = 12/03/79)

* M U L T I P L E R E G R E S S I O N * * * * * * * * * * * * * *

DEPENDENT VARIABLE.. X20

COEFFICIENTS AND CONFIDENCE INTERVALS.

VARIABLE	B	STD ERROR B	T	95.0 PCT CONFIDENCE INTERVAL	
X2	8.2056618	17.346404	.47304686	-26.671620	, 43.082944
X30	-.35382193	.25552978	-1.3846603	-.86759899	, .15995513
X34	59.533507	207.10304	.28745839	-356.87507	, 475.94208
X5	.99001459	.83450842	1.1863446	-.68787706	, 2.6679062
X11	.12680075	.31751393	.39935493	-.51160379	, .76520529
X27	.49276704	.61187387	.80534088	-.73746784	, 1.7230219
X28	.64676379E-01	.13817898	.46806236	-.21315108	, .34250384
X4	-.33287310	.39887747	-.83452470	-1.1348700	, .46912381
X32	531.35247	667.96390	.79548081	-811.67898	, 1874.3839
X14	1102.0708	280.30031	3.9317500	538.48920	, 1665.6523
X22	-.13099197	.22804494	-.57441295	-.58950705	, .32752312
X15	.36000615	.17946696	2.0059745	-.83636371E-03	, .72084667
X17	.60315963	.43885436	1.3743959	-.27921622	, 1.4855355
X8	-1.3736140	2.5219780	-.54465739	-6.4443906	, 3.6971627
X33	.55957127E-01	.28894330	.19366147	-.52500172	, .63691598
X29	-.13646127	.43869986	-.31105839	-1.0185265	, .74560392
X21	.13008302	.15751183	.82586193	-.18661574	, .44678178
X10	2.6000225	1.6961873	1.5328628	-.81039070	, 6.0104356
X31	-.36370529	.46191312	-.78738895	-1.2924439	, .56503330
X25	-.75005236	2.6271934	-.28549567	-6.0323788	, 4.5322741
X26	-.14280299	.60250900	-.23701387	-1.3542285	, 1.0686226
X23	-.34773058	1.1390687	-.30527621	-2.6379817	, 1.9425206
CONSTANT	-10.438349	1.4928598	-6.9921834	-13.439945	, -7.4367536

REGRESSION .OGNEWDATA 12/03/79 15.30.20. PAGE 9

FILE NONAME (CREATION DATE = 12/03/79)

* M U L T I P L E   R E G R E S S I O N *

DEPENDENT VARIABLE.. X20

S U M M A R Y   T A B L E

STEP	VARIABLE ENTERED	VARIABLE REMOVED	F TO ENTER OR REMOVE	SIGNIFICANCE	MULTIPLE R	R SQUARE	R SQUARE CHANGE	SIMPLE R	OVERALL F	SIGNIFICANCE
1	X2		.22377	.638	.28705	.08240	.08240	.28705	7.73752	.000
	X30		1.91728	.173	.28923	.08365	.00125	.02840		
	X34		.08263	.775	.53750	.28890	.20525	.44536		
	X5		1.40741	.241	.54491	.29692	.00802	-.04738		
	X11		.13948	.691	.54807	.30038	.00346	.07241		
	X27		.64857	.425	.56315	.31714	.01676	.18454		
	X28		.21908	.642	.61487	.37807	.06093	.37205		
	X4		.69643	.408	.61987	.38423	.00617	.00588		
	X32		.63279	.430	.65772	.43259	.04836	.08281		
	X14		15.45866	.000	.68829	.47374	.04115	.30441		
	X22		.32995	.568	.70059	.49082	.01708	.15652		
	X15		4.02393	.051	.83196	.69215	.20133	.72005		
	X17		1.89896	.176	.85729	.73494	.04279	.65839		
	X8		.29665	.589	.87302	.76216	.02722	-.02685		
	X33		.03730	.847	.87302	.76216	.00000	.48474		
	X29		.09675	.757	.87308	.76227	.00011	.49228		
	X21		.68205	.413	.87472	.76514	.00286	.18672		
	X10		2.34967	.132	.88114	.77641	.01127	-.03536		
	X31		.61998	.435	.88270	.77915	.00274	.07299		
	X25		.08151	.776	.88283	.77939	.00024	.28453		
	X26		.05616	.814	.88296	.77962	.00022	.07141		
	X23		.09319	.761	.88320	.78084	.00043	.28635		

APPENDIX F

SELECTED LEAPS AND BOUNDS OUTPUT

```
MAXIMUM NUMBER OF VARIABLES = 26
THERE ARE  25 INDEPENDENT VARIABLES
MAXIMUM NUMBER OF OBSERVATIONS = 100
NUMBER OF OBSERVATIONS =  63
        INPUT MATRIX
 1
 443.0000000     17.70000000     410.0000000     200.0000000     85.00000000     0.
 0.              0.              0.              0.              0.              0.
 0.              443.0000000     17.70000000     410.0000000     200.0000000     0.
 0.              0.              0.              0.              0.              1.000000000
 0.              .5934419548
 2
 936.0000000     36.00000000     2176.000000     175.0000000     73.00000000     0.
 0.              0.              0.              0.              0.              0.
 0.              993.0000000     36.00000000     2176.000000     175.0000000     0.
 0.              0.              0.              0.              0.              1.000000000
 0.              .6930947274
 3
 1444.000000     40.00000000     491.0000000     212.0000000     75.00000000     25.00000000
 0.              0.              0.              0.              0.              0.
 0.              1444.000000     40.00000000     491.0000000     212.0000000     1444.000000
 40.00000000     491.0000000     212.0000000     75.00000000     25.00000000     1.000000000
 1.000000000     .8414589651
 4
 1676.000000     30.00000000     78.00000000     820.0000000     24.00000000     0.
 0.              0.              0.              0.              0.              0.
 0.              1676.000000     30.00000000     78.00000000     820.0000000     0.
 0.              0.              0.              0.              0.              1.000000000
 0.              6.067775943
 5
 1473.000000     40.00000000     689.0000000     77.00000000     74.00000000     25.00000000
 0.              0.              0.              0.              0.              0.
 0.              1473.000000     40.00000000     689.0000000     77.00000000     1473.000000
 40.00000000     689.0000000     77.00000000     74.00000000     25.00000000     1.000000000
 1.000000000     1.115773118
 6
 1276.000000     36.50000000     758.0000000     289.0000000     85.00000000     0.
 0.              0.              0.              0.              0.              0.
 0.              1276.000000     36.50000000     758.0000000     289.0000000     0.
 0.              0.              0.              0.              0.              1.000000000
 0.              .[illegible]81116
 7
 91.00000000     4.000000000     12.00000000     20.00000000     150.0000000     0.
 0.              0.              0.              0.              0.              0.
 0.              91.00000000     4.000000000     12.00000000     20.00000000     0.
 0.              0.              0.              0.              0.              1.000000000
 0.              .5411953515E-01
 8
 133.0000000     1.250000000     84.00000000     87.00000000     97.00000000     0.
 0.              0.              0.              0.              0.              0.
 0.              133.0000000     1.250000000     84.00000000     87.00000000     0.
 0.              0.              0.              0.              0.              1.000000000
 0.              [illegible]
```

REGRESSIONS WITH 15 VARIABLE(S) (R-SQUARED)

CRITERION	VARIABLES
.762794E+02	1 2 4 5 9 11 14 16 18 19 20 21 23 24 25
.766551E+02	1 2 4 5 9 11 14 16 18 19 20 21 22 23 24

REGRESSIONS WITH 16 VARIABLE(S) (R-SQUARED)

CRITERION	VARIABLES
.783593E+02	1 2 4 5 6 9 11 14 16 18 19 20 21 23 24 25
.777834E+02	1 2 4 6 7 9 11 12 14 16 18 19 20 21 23 25

REGRESSIONS WITH 17 VARIABLE(S) (R-SQUARED)

CRITERION	VARIABLES
.795391E+02	1 2 4 6 7 9 11 12 14 16 17 18 19 20 21 23 25
.787742E+02	1 2 4 5 6 7 9 11 14 16 18 19 20 21 23 24 25

REGRESSIONS WITH 18 VARIABLE(S) (R-SQUARED)

CRITERION	VARIABLES
.799449E+02	1 2 4 6 7 8 9 11 12 14 16 17 18 19 20 21 23 25
.799159E+02	1 2 4 5 6 7 9 11 12 14 16 18 19 20 21 23 24 25

REGRESSIONS WITH 19 VARIABLE(S) (R-SQUARED)

CRITERION	VARIABLES
.801552E+02	1 2 4 5 6 7 8 9 11 12 14 16 17 18 19 20 21 23 25
.801273E+02	1 2 4 5 6 7 9 11 12 14 16 17 18 19 20 21 23 24 25

REGRESSIONS WITH 20 VARIABLE(S) (R-SQUARED)

CRITERION	VARIABLES
.803551E+02	1 2 4 5 6 7 8 9 11 12 14 16 17 18 19 20 21 23 24 25
.802144E+02	1 2 4 5 6 7 9 11 12 14 16 17 18 19 20 21 22 23 24 25

REGRESSIONS WITH 21 VARIABLE(S) (R-SQUARED)

CRITERION	VARIABLES
.804339E+02	1 2 4 5 6 7 8 9 11 12 14 15 17 18 19 20 21 22 23 24 25
.797070E+02	1 2 3 4 6 7 8 9 10 11 12 13 15 16 17 18 19 20 21 23 25

```
REGRESSIONS WITH  15 VARIABLE(S) (R-SQUARED)

      CRITERION         VARIABLES
       .762734E+02        1  2  4  5  9 11 14 16 18 19 20 21 23 24 25
       .760551E+02        1  2  4  5  9 11 14 16 18 19 20 21 22 23 24

REGRESSIONS WITH  16 VARIABLE(S) (R-SQUARED)

      CRITERION         VARIABLES
       .753593E+02        1  2  4  5  6  9 11 14 16 18 19 20 21 23 24 25
       .777834E+02        1  2  4  6  7  9 11 12 14 16 18 19 20 21 23 25

REGRESSIONS WITH  17 VARIABLE(S) (R-SQUARED)

      CRITERION         VARIABLES
       .795391E+02        1  2  4  6  7  9 11 12 14 16 17 18 19 20 21 23 25
       .787742E+02        1  2  4  5  6  7  9 11 14 16 18 19 20 21 23 24 25

REGRESSIONS WITH  18 VARIABLE(S) (R-SQUARED)

      CRITERION         VARIABLES
       .799449E+02        1  2  4  6  7  8  9 11 12 14 16 17 18 19 20 21 23 25
       .799059E+02        1  2  4  5  6  7  9 11 12 14 16 18 19 20 21 23 24 25

REGRESSIONS WITH  19 VARIABLE(S) (R-SQUARED)

      CRITERION         VARIABLES
       .801552E+02        1  2  4  5  6  7  8  9 11 12 14 16 17 18 19 20 21 23 25
       .801273E+02        1  2  4  5  6  7  9 11 12 14 16 17 18 19 20 21 23 24 25

REGRESSIONS WITH  20 VARIABLE(S) (R-SQUARED)

      CRITERION         VARIABLES
       .803551E+02        1  2  4  5  6  7  8  9 11 12 14 16 17 18 19 20 21 23 24 25
       .802144E+02        1  2  4  5  6  7  9 11 12 14 16 17 18 19 20 21 22 23 24 25

REGRESSIONS WITH  21 VARIABLE(S) (R-SQUARED)

      CRITERION         VARIABLES
       .804339E+02        1  2  4  5  6  7  8  9 11 12 14 15 17 18 19 20 21 22 23 24 25
       .797070E+02        1  2  3  4  6  7  8  9 10 11 12 13 15 16 17 18 19 20 21 23 25
```

BEST REGRESSIONS WITH 11 VARIABLE(S) (R-SQUARED)

VARIABLE	COEFFICIENT	PARTIAL F	ALPHA
1	-.153111E-02	.107769E+02	.185974E-02
2	.721023E-01	.173720E+02	.119046E-03
4	.142447E-02	.958091E+01	.319066E-02
9	.377547E-01	.562302E+01	.130170E-01
11	-.199900E-02	.534911E+01	.243546E-01
14	.116613E-02	.217056E+02	.231227E-04
16	-.100916E-02	.133714E+02	.604651E-03
18	.902775E-03	.334731E+01	.731634E-01
19	-.443078E-01	.495599E+01	.299540E-01
20	.139595E-02	.152265E+02	.248600E-03
21	-.396217E-02	.183168E+02	.824527E-04

BEST REGRESSIONS WITH 12 VARIABLE(S) (R-SQUARED)

VARIABLE	COEFFICIENT	PARTIAL F	ALPHA
1	-.173056E-02	.124637E+02	.901575E-03
2	.752113E-01	.187133E+02	.729292E-04
4	.119269E-02	.584349E+01	.193233E-01
5	-.513852E-02	.157886E+01	.214764E+00
9	.390592E-01	.713294E+01	.101646E-01
11	-.211154E-02	.595583E+01	.181537E-01
14	.120856E-02	.235294E+02	.124211E-04
16	-.968316E-03	.103068E+02	.231752E-02
18	.103725E-02	.426501E+01	.440850E-01
19	-.472216E-01	.560371E+01	.213155E-01
20	.123163E-02	.107408E+02	.190963E-02
21	-.291218E-02	.162957E+02	.185946E-03

BEST REGRESSIONS WITH 13 VARIABLE(S) (R-SQUARED)

VARIABLE	COEFFICIENT	PARTIAL F	ALPHA
1	-.157972E-02	.105328E+02	.200767E-02
2	.737949E-01	.189438E+02	.532766E-04
4	.116113E-02	.603669E+01	.175995E-01
5	-.179543E-01	.531797E+01	.253781E-01
9	.444804E-01	.937662E+01	.356350E-02
11	-.259589E-02	.671737E+01	.482552E-02
14	.927114E-03	.889369E+01	.449102E-02
16	-.385162E-03	.838763E+01	.563067E-02
18	.126491E-02	.631222E+01	.153326E-01
19	-.517025E-01	.703423E+01	.107450E-01
20	.995016E-03	.465894E+01	.322304E-01
21	-.270753E-02	.145058E+02	.369684E-03
24	.151363E+01	.367756E+01	.609877E-01

```
REGRESSIONS WITH  15 VARIABLE(S) (ADJUSTED R-SQUARED)

      CRITERION          VARIABLES
       .696983E+02        1  2  4  5  6  9 11 14 18 19 20 21 23 24 25
       .694739E+02        1  2  4  5  6 10 11 16 18 19 20 21 23 24 25

REGRESSIONS WITH  16 VARIABLE(S) (ADJUSTED R-SQUARED)

      CRITERION          VARIABLES
       .704565E+02        1  2  4  5  6  9 11 14 16 18 19 20 21 23 24 25
       .690313E+02        1  2  4  5  6 10 11 14 16 18 19 20 21 23 24 25

REGRESSIONS WITH  17 VARIABLE(S) (ADJUSTED R-SQUARED)

      CRITERION          VARIABLES
       .710091E+02        1  2  4  6  7  9 11 12 14 16 17 18 19 20 21 23 25
       .707556E+02        1  2  4  5  6  7  9 11 14 16 18 19 20 21 23 24 25

REGRESSIONS WITH  18 VARIABLE(S) (ADJUSTED R-SQUARED)

      CRITERION          VARIABLES
       .716355E+02        1  2  4  5  6  7  9 11 12 14 16 18 19 20 21 23 24 25
       .715133E+02        1  2  4  5  6  7  9 11 12 14 16 17 18 19 20 21 23 25

REGRESSIONS WITH  19 VARIABLE(S) (ADJUSTED R-SQUARED)

      CRITERION          VARIABLES
       .713453E+02        1  2  4  5  6  7  9 11 12 14 16 17 18 19 20 21 23 24 25
       .712384E+02        1  2  4  5  6  7  9 11 12 14 16 18 19 20 21 22 23 24 25

REGRESSIONS WITH  20 VARIABLE(S) (ADJUSTED R-SQUARED)

      CRITERION          VARIABLES
       .707927E+02        1  2  4  5  6  7  9 11 12 14 16 17 18 19 20 21 22 23 24 25
       .698834E+02        1  2  3  4  5  6 10 11 13 14 15 16 18 19 20 21 22 23 24 25

REGRESSIONS WITH  21 VARIABLE(S) (ADJUSTED R-SQUARED)

      CRITERION          VARIABLES
       .704159E+02        1  2  4  5  6  7  8  9 11 12 14 16 17 18 19 20 21 22 23 24 25
       .693131E+02        1  2  3  4  6  7  8  9 10 11 12 13 15 16 17 18 19 20 21 23 25
```

BEST REGRESSIONS WITH 17 VARIABLE(S) (ADJUSTED R-SQUARED)

VARIABLE	COEFFICIENT	PARTIAL F	ALPHA
1	-.155699E-02	.117394E+02	.129444E-02
2	.779107E-01	.227635E+02	.196318E-04
4	.105464E-02	.552395E+01	.231433E-01
5	.361793E-01	.875757E+01	.490294E-02
7	.201123E+01	.758031E+01	.640010E-02
9	.710891E-01	.144932E+02	.422849E-03
11	-.500173E-02	.180398E+02	.108336E-03
12	-.267022E-01	.762132E+01	.631715E-02
14	.870194E-03	.977180E+01	.339106E-02
16	-.120070E-02	.173296E+02	.115912E-03
17	.143445E-02	.385836E+01	.550770E-01
18	.204243E-02	.122373E+02	.104061E-02
19	-.112440E+00	.134675E+02	.640615E-03
20	.214320E-02	.151595E+02	.323059E-03
21	-.402165E-02	.220968E+02	.247408E-04
23	.153040E+00	.840314E+01	.570573E-02
25	-.547019E+01	.777109E+01	.774979E-02

VITA

Joseph Richard Cafarella, Jr. was born on November 27, 1956 in Moses Lake, Washington. He graduated from Northern Burlington Regional High School in New Jersey in 1974 and attended the Virginia Military Institute where he graduated in 1978 with a Bachelor or Science degree and a double major in Mathematics and Civil Engineering and a reserve commission in the United State Air Force. In June of 1978 he entered the Air Force Institute of Technology.

Permanent Address: 311 W. Michigan
Electra, Texas 76360

This thesis was typed by Donna K. Beam.

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: AFIT/GOR/MA/79D-2 ✓

2. GOVT ACCESSION NO.

3. RECIPIENT'S CATALOG NUMBER

4. TITLE (and Subtitle): CROSS VALIDATION OF SELECTION OF VARIABLES IN MULTIPLE REGRESSION

5. TYPE OF REPORT & PERIOD COVERED: M.S. Thesis

6. PERFORMING ORG. REPORT NUMBER

7. AUTHOR(s): Joseph R. Cafarella, 2Lt, USAF

8. CONTRACT OR GRANT NUMBER(s)

9. PERFORMING ORGANIZATION NAME AND ADDRESS: AFIT/EN, Wright-Patterson AFB, Ohio 45433

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS

11. CONTROLLING OFFICE NAME AND ADDRESS: Air Force Institute of Technology (AFIT/EN), Wright-Patterson AFB, Ohio 45433

12. REPORT DATE: December 1979

13. NUMBER OF PAGES: 124

14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)

15. SECURITY CLASS. (of this report): Unclassified

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

JOSEPH P. HIPPS, MAJ, USAF
DIRECTOR OF PUBLIC AFFAIRS

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Cross Validation
Selection of Variables
Multiple Regression

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Techniques and criterion for selection of the "best" subset of variables to be used in a regression model are reviewed.

A model was developed using the Automatic Interaction Detection (AID) algorithm as a pre-screening device for locating those variables most important to the regression including interaction terms.

Five previous models including the one developed by AID and one developed by Westinghouse on avionic characteristic data are used in cross validation experiments to determine the predictive power of these models on a new set of

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

Block 20 con't

data points using the same set of variables. A cross validation R^2 value is discussed as a criterion for choosing between competing models.